



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

기계학습을 이용한

Aspect-Based Sentiment Analysis 기반 전기차

요소별 사용자 감성 분석 및 예측 모델링

Aspect-Level Analysis and Predictive Modeling for Electric
Vehicle Based on Aspect-Based Sentiment Analysis Using
Machine Learning

2020 년 8 월

서울대학교 대학원

산업공학과

정 세 준

기계학습을 이용한 Aspect-Based Sentiment Analysis 기반 전기차 요소별 사용자 감성 분석 및 예측 모델링

Aspect-Level Analysis and Predictive Modeling for Electric Vehicle Based on Aspect-Based Sentiment Analysis Using Machine Learning

지도교수 윤 명 환

이 논문을 공학석사 학위논문으로 제출함

2020 년 8 월

서울대학교 대학원

산업공학과

정 세 준

정세준의 공학석사 학위논문을 인준함

2020 년 8 월

위 원 장 _____ 조 성 준

부위원장 _____ 윤 명 환

위 원 _____ 박 건 수



초록

본 논문에서는 전기차를 대상으로 기계학습을 이용한 Aspect-Based Sentiment Analysis(ABSA) 기반 사용자 리뷰 분석을 통해, 차량의 주요 요소(Aspect)인 부품(Components) 및 특징(Attributes)을 추출하고, 추출된 각 요소에 대한 사용자 감성 예측 모델링 기반의 UX 분석 프레임워크(Framework)를 구현하여 기존의 인터뷰 및 설문조사와 유사한 수준의 사용자 의견을 얻는 것을 주요 목표로 한다. 이 과정에서 수반되는 데이터 불균형(Data Imbalance) 문제를 오버샘플링(Oversampling)을 통해 극복하고, 사용자 리뷰 부족 문제 극복을 위해 레이블이 없는(Non-label) 데이터를 활용하는 방법을 제안한다. 더불어 추출된 Aspect에 대한 차량 세부 스펙과 사용자 감성 간의 관계성 확인을 통해 감성에 영향을 주는 요소(Contributing Factor)를 찾는다.

연구 방법은 ABSA의 큰 틀을 활용하며, 크게 데이터 수집, 전처리 및 Feature 생성, 요소 추출(Asspect Extraction) 및 감성 분석(Sentiment Analysis)을 위한 모델링, 그리고 요소 별 사용자 감성 분석 순서로 진행하였다. 데이터 수집은 대표적인 자동차 포럼에서 사용자 만족도가 5점 척도로 평가된 Label 데이터 총 5,065개를 수집하였고, 데이터 부족 문제를 극복하고자 Youtube.com에서 Non-label 데이터를 약 21만개 수집하였으며 이 중 User Experience 관련 어휘가 포함된 리뷰로 한정하여 총 6,488개를 선별하였다. 이후 수집 데이터의 전처리 및 분산 표현(Distributed Representation)을 통한 효과적인 임베딩 과정을 거쳐 특징(Feature)을 생성하였다. 분석은 크게 두 가지 줄기로써, 요소 추출(Asspect Extraction)과 감성 분석(Sentiment Analysis)로 나뉜다. 요소 탐지를 위해 비지도적 방법(Unsupervised Method)이자 추출적 접근 방법(Extractive Approach)으로써, TextRank와 Naïve Method를 활용하였다. 그 다음 지도학습(Supervised Learning) 기반의 문장 감성 분류 모델을 구현하고자, Label이 있는 리뷰 텍스트 서두의 한 두문장으로 구성된 절단된 텍스트를 학습시킨 모델을

구축하였고, 준지도학습을 통해 더 나은 성능의 모델을 구현하고자 하였다. 이를 바탕으로 선정된 Aspect가 포함된 문장에 대한 감성 분석을 실시함으로써 요소별 감성 분석을 진행하고, 더불어 사용자 감성에 영향력 있는 차량 세부 스펙을 찾아 Contributing Factor를 발굴하고자 하였다.

연구 결과로써, 요소 추출(Aspect Extraction)로는 총 16개 카테고리의 주요 Aspects(8개의 주요 전기차 구성 요소와 8개의 주요 Human Factor 특성)가 추출되었는데, 이 중 사용자는 “Acceleration / Room / Interior / Power / Safety / Ergonomics / Price / Power”에 대해 긍정적이며, “Seat / Battery / Charge / Noise / Winter / Ice”에 대해 다소 부정적임을 확인하였다. 감성 분석(Sentiment Analysis)에서는 CNN 모델이 리뷰 단위 감성 분류에 있어 가장 높은 성능을 보였다. 따라서 CNN을 활용한 준지도학습(Semi-Supervised Learning)을 통해 Non-Label Data 중 80% 이상의 분류 확률이 높은 데이터 위주로 Pseudo Label을 부여하였고, 이를 포함한 전체 데이터를 재학습을 거치는 방법으로 모델의 성능 향상을 확인하였다. 또한 추출된 요소가 포함된 문장 단위 감성 분류에 대하여, 기계학습 모델 기반으로 결과와 Lexicon 기반 감성 분류 결과 간 17개 Aspect 중 14개가 예측 방향성이 일치함을 확인함으로써, 기계학습 기반 감성 분류 모델의 타당성을 간접적으로 확인하였다. 마지막으로 샘플 검증을 통해 본 연구에서 학습된 딥러닝 모델의 높은 분류 정확도를 확인하였는데, 딥러닝 모델이 단어 의미 이상으로 문장 문맥을 파악하여 긍정/부정 분류하였음을 확인하였다.

결론적으로 Aspect 기반의 문장단위 분석을 통해 보다 더 다양한 토픽과 편향되지 않은 의견을 추출할 수 있음을 보였다. 더불어 리뷰 데이터를 Over-sampling을 하여 Data Imbalance 문제를 접근함으로써 온라인 리뷰의 긍정 편향성을 극복하고, Semi-Supervised Learning을 통한 Non-Label Data 활용 방법을 통해 사용자 평가가 많이 부족한 제품에 대해 보다 효과적인 UX 분석 프레임워크를 제안하였다.

주요어: Aspect Based Sentiment Analysis, 사용자 경험, 기계학습

학번: 2018-23341

목차

초록	ii
목차	iv
표 목차	vi
그림 목차	vii
제 1 장 서론	1
1.1 연구 배경.....	1
1.2 연구 대상.....	3
제 2 장 연구 목표	5
2.1 연구 목표	5
2.2 선행 연구	7
2.2.1 Aspect-Based Sentiment Analysis	7
2.2.2 요소 추출(Asspect Extraction)	9
2.2.3 감성 분석(Sentiment Analysis)	12
제 3 장 연구 방법	18
3.1 데이터 수집.....	19
3.1.1 Label 데이터 수집	19
3.1.2 Non-Label 데이터 수집	19
3.1.3 데이터 분포	20

3.2	데이터 전처리	21
3.3	Aspect-Based Sentiment Analysis 수행	24
3.3.1	요소 추출(Asspect Extraction)	25
3.3.2	감성 분석(Sentiment Analysis)	27
3.3.3	A Framework for UX Analysis	33
제 4 장	연구 결과	34
4.1	요소 추출(Asspect Extraction) 결과	34
4.2	감성 분석(Sentiment Analysis)을 위한 모델링 결과	36
4.2.1	기계학습 기반 감성 분류 결과에 대한 모델 별 성능 비교	36
4.2.2	준지도학습(Semi Supervised Learning) 실험 결과	38
4.3	Aspect Based Sentiment Analysis 결과	39
4.3.1	기계학습 모델 기반 ABSA 결과	39
4.3.2	Lexicon 기반 ABSA 결과	40
4.4	사용자 긍정/부정 경험에 영향을 미치는 Contributing Factor	42
제 5 장	결론	44
5.1	결론(Conclusion)	44
5.2	연구 기여(Contribution)	46
5.3	한계점(Limitation)	46
Appendix		47
참고문헌		49
Abstract		55
감사의 글		57

표 목차

Table 4.1	요소 추출 결과	35
Table 4.2	감성 분석 실험 결과	37
Table 4.3	딥러닝 모델의 문장 분류 성능 실험	40
Table 4.4	Lexicon Based 모델과 ML 모델간 문장 긍정/부정 분류 결과 비교	42
Table 4.5	“Room”에 대한 사용자 감성 극성과 상세 스펙간 회귀분석 결과	43

그림 목차

Figure 1.1	연구 배경 및 핵심 이슈	3
Figure 2.1	연구 주요 목표와 하위 목표 및 최종 산출물	6
Figure 2.2	CNN for Sentence Classification	13
Figure 2.3	Attention-based bidirectional LSTM for Classification	14
Figure 2.4	Hierarchical Attention Network for Text Classification	15
Figure 2.5	Self-Training	16
Figure 3.1	ABSA 기반 사용자 경험 분석 Framework	18
Figure 3.2	수집된 데이터의 Lexicon 기반 감성 분포	21
Figure 3.3	전처리 결과 I	23
Figure 3.4	전처리 결과 II	23
Figure 3.5	Borderline-SMOTE	30
Figure 3.6	A Brand-new Framework: Analysis of User Experience	34
Figure 4.1	준지도학습 실험 결과	38
Figure 4.2	기계학습 모델(CNN) 기반 ABSA 결과	39
Figure 4.3	Lexicon 기반 ABSA 결과 I	41
Figure 4.4	Lexicon 기반 ABSA 결과 II	41
Figure 4.5	Aspect의 차량 별 실제 치수와 사용자 감성 간 상관 관계 예시	43
Figure 5.1	Aspect-wise 분석과 Review-wise 분석 비교	44
Figure 5.2	전기차 긍정/부정 요소 및 도출된 사용자 니즈	45

제 1 장 서론

1.1 연구 배경

사용자 중심 디자인(User Centered Design)을 위해 사용자를 이해하고 사용 맥락(Context of User)을 파악하는 단계에서, 설문조사(User Survey), 사용자 관찰(Observation), 카드 소팅(Card Sorting), 그리고 인터뷰(Focus Group Interview)는 연구자의 목표에 직접적으로 부합하는 답변을 구체적으로 얻을 수 있는 장점이 있어서 전통적으로 가장 많이 선호되고 있는 방법들이다. 다만 본 방법들은 많은 실행 인력과 비용이 요구될 뿐만 아니라 직접 대면이 필요한 방법이므로, 최근 COVID-19로 인하여 사회적으로 언택트(Untact)가 요구되는 시점에서는 사용자 조사 및 실험 자체가 진행되기 어려울 수도 있다. 그리고 제품 수명 주기(Product Life Cycle) 상 성장 초기(Early Growth) 단계에 해당하는 제품의 경우, 제품을 경험한 다수의 사용자들을 모집하는 데 큰 어려움이 발생한다. 따라서 텍스트마이닝을 통한 분석이 보다 더 효과적일 수 있으며, Web, App, SNS 등의 디지털 플랫폼 상에서 대량으로 실시간 발생하는 사용자 의견을 빠르게 파악하고 대응할 수 있는 장점이 있기에, 본 연구는 텍스트 마이닝 기반의 효과적인 사용자 경험(User Experience)을 분석 방법론을 정립하여 인터뷰 및 설문조사와 유사한 수준의 사용자 의견을 얻을 수 있는 방법을 고안해보고자 한다.

본 연구에서 사용자 경험 분석을 위해 활용한 기반 프레임워크는 Aspect-Based Sentiment Analysis(ABSA)이다. ABSA의 목적은 사용자가 생성한 리뷰 텍스트 내에서 제품 및 서비스의 세부 개체(Entity)와 특징(Attributes)에 대한 의견을 얻는 것으로, 이는 사용자가 작성하는 평점(User Rating) 또는 일반적인 리뷰 단위 감성 분석(Sentiment Analysis)로부터 얻을 수 없는 정보이다. 즉, ABSA는 제품에 대한 대량의 사용자 리뷰 문서로부터 특정 구성 부품(Component) 및 개체(Entity)와 사용자와 제품 간의 인터랙션 특성(Attribute)에 관한 정보력 있는 단어(Informative

Word)를 추출함으로써 제품 요소(Aspect)를 얻고, 각 요소에 대한 사용자 감성을 파악할 수 있다.

본 논문에서 제안하는 기계학습 기반의 Aspect-Based Sentiment Analysis 방법론은 특히 사용자 리뷰가 매우 부족한 제품의 사용자 경험을 분석하여 주요 요소(Aspect)와 요소 별 사용자 감성 및 Contributing Factor를 파악할 수 있기에, 향후 수요가 급격히 증가할 수 있는 차세대 제품에 대해서도 사용자 인사이트를 얻을 수 있다는 점에서 관련 분야에 공헌하고자 한다.

본 연구에서는 제품 “전체” 보다는 제품의 각 “요소(Aspect)”에 대한 사용자의 감성이 각각 어떠한 지 파악해야 다양한 토픽과 긍정/부정 한쪽으로 편향되지 않은 사용자 의견을 얻을 수 있을 것이라고 가정했다. 그 이유는 온라인 상에는 긍정적인 평점의 리뷰가 압도적으로 많다. 이는 일종의 편향 현상으로 긍정적인 경험을 얻은 사용자만 긍정적인 의견을 남긴다는 사용자 행동 성향 때문이며, 흔히 알파벳 J자 모양(J-shape) 형태로 리뷰가 분포하고, 이때 긍정 리뷰가 다수이다 [1]. 하지만 사용자들이 전반적인 제품에 대한 긍정적인 리뷰를 남기더라도, 제품의 모든 요소에 대해서 모두 긍정적이지 않을 것이다. 예를 들어 사용자가 전기차를 구입한 경우, 전기차가 전반적으로 마음에 들어서 긍정적인 리뷰 평점을 남겼을 지라도, 전기차의 차량 시트, 핸들과 같은 일부 구성 부품이나 가속감, 승차감, 안전성과 같은 사용 특성에 대해서는 부정적일 수 있으며 이는 사용자가 작성한 리뷰에 표현되어 있다. 따라서 텍스트 내에 내재된 제품 특정 요소에 대한 사용자의 부정적인 감성을 추출해야 결국 사용자가 어느 부분의 개선을 원하는 지 발굴할 수 있기에, 본 연구에서는 요소(Aspect)별 감성 분석에 주안점을 두었다. 더불어 사용자의 감성에 영향력 있는 제품의 스펙을 발굴하고 상관관계를 확인함으로써 사용자 중심 디자인을 위한 인사이트를 얻고자 한다. 정리하자면, 제품의 주요 요소(Aspect)를 추출하여 각각에 대한 사용자의 긍정/부정 감성을 확인하고, 제품 요소의 상세 스펙과 사용자가 느끼는 긍정/부정 감성 극성 간의 상관관계를 분석할 수 있는 하나의 새로운 텍스트 마이닝 기반의 UX 평가

Framework의 구축하고자 본 연구를 진행하게 되었다.



Figure 1.1 연구 배경 및 핵심 이슈

1.2 연구 대상

본 연구에서는 대상 제품을 “전기차”로 선정하였다. 이는 최근 정부의 그린 뉴딜과 같은 친환경 정책에 따라 향후 전기차 수요가 가파르게 증가할 것으로 예상되는 만큼, 사용자 친화적인 제품 설계를 위하여 현재까지의 전기차에 대한 사용자 니즈와 부정 요소 발굴을 통해 관련 부품, 디자인, 서비스 등의 개선이 필요한 시점이기 때문이다.

전기차는 배출 가스 규제 강화 및 친환경 정책과 맞물려 현재 기업들이 전기차 중심으로 사업을 빠르게 재편하고 있기에 더욱 높은 시장성이 기대된다. 먼저 국내 시장에서 전기차 보급이 점차 늘어날 것으로 예상되는데, 2013년 산업수요 대비 비중이 0%였던 전기차는 올해 2.4%까지 비중을 늘릴 예정이고, “2030년 미래차 산업 발전전략”에 따르면 2030년에는 연간 신차판매 중 전기·수소차 비중을 24.4%까지 늘릴 계획이라고 발표한 바 있다 [2].

이와 같이 전기차가 빠른 기간 내에 시장을 형성하고 규모를 키워 나갈 수 있었던 원동력은 무엇보다도 국가 보조금 및 세제혜택 등 정책적 지원을 초기 구매 비용 절감과 전기차의 저렴한 유지 비용에 따른 총 소유 비용(Total Cost of Ownership) 절감 측면에서 소비자의 구미를 당긴 요인이 크다. 정책적 지원은 점차

줄어들고 있지만, 2025년부터는 배터리 가격 하락으로 보조금 없이도 가격 경쟁력을 확보, 내연기관차와의 경쟁에서 경량차는 2030년부터, 중형차는 2020년 후반부터 전기차로의 전환이 가속될 것으로 보고 있다.

전 세계적으로도 전기차의 빠른 확산이 예상되는데, 블룸버그 뉴에너지 파이낸스(BNEF)에서 발간한 Electric Vehicle Outlook 2017 [3]의 Executive Summary에 따르면, 2025년까지 상대적으로 낮은 전기자동차 판매량은 2025년에서 2030년 사이에 급증하여 2040년 신차 판매량의 54%, 전 세계 자동차의 33%를 점유할 것으로 전망하고 있다. 또한 국내 에너지연구원에 따르면 주요 선진국들은 전기차를 국제 환경규제 대응 및 자국 자동차산업의 경쟁력 향상을 위한 핵심 기술로 인식하여 전기차 보급정책을 강력히 추진하고 있다. 실제로 현재 유럽연합이 추진 중인 배출가스 규제 강화 조치는 각 업체가 판매하는 차량의 평균 이산화탄소(CO₂) 배출량 기준을 km당 130g에서 95g으로 강화하는 게 핵심인데, 본 규제는 내년부터 적용된다. EU는 파리기후변화협약에 따라 차량당 이산화탄소 배출 허용량을 현 km당 130g에서 2050년까지 km당 10g으로 단계적으로 줄여 나간다는 계획을 밝히고 전기차 시장 육성을 유도해왔고, 각국 정부도 이에 따라 규제를 강화하는 추세이다. 따라서 유럽연합(EU)의 배출가스 규제가 대폭 강화되는 내년을 기점으로 전기차 보급이 이전과는 다른 양상을 보일 것이라고 전망한다. 각국의 환경 규제를 피하려는 완성차 업체들이 속속 새로운 전기차 차종을 출시하면서 경쟁 구도가 형성되고 있고, 소비자의 선택권이 더욱 확대됨에 따라 성장세가 더욱더 가팔라질 것으로 예상하기 때문이다.

제 2 장 연구 목표

2.1 연구 목표

본 연구의 주요 목표는 전기차와 같은 성장 초기(Early Growth) 단계의 향후 수요가 폭발적인 증가가 예상되는 제품에 대하여 텍스트마이닝(Text-mining) 기반의 사용자 경험 분석을 통해 제품의 주요 구성 요소(Component) 및 사용자-차량 간 인터랙션(Human-Vehicle Interaction) 특성을 파악하여 전기차의 주요 요소(Aspect)를 발굴하고, 각 요소(Aspect)에 대한 사용자가 느끼는 감성을 파악하는 것이다. 이를 통해 각 요소에 대한 사용자 요구사항 및 니즈를 추출하고, 해당 감성에 영향력 있는 요인(Contributing Factor)를 파악하면서 사용자 감성에 영향을 미치는 주요 차량 스펙(Specification)을 확인한다. 이 과정에서 수반되는 사용자 리뷰 부족문제를 극복하기 위해 Non-Label 데이터를 함께 활용하는 방법을 도출하여 전기차와 같이 사용자 리뷰가 부족한 차세대 제품에 대한 더 다양한 토픽을 얻고, 또한 데이터 불균형(Data Imbalance) 문제를 극복하여 보다 견고한 감성 예측 모델을 만들고자 한다.

기본적으로 텍스트마이닝을 통해 사용자가 작성한 리뷰에서 유의미한 정보를 발굴하고자 하였다. 텍스트마이닝은 관계 테이블과 같이 정형화된 자료에서 어떤 지식을 얻어내기보다는 자연어처리(Natural Language Processing) 기반으로 대량의 비정형 텍스트 문서로부터 유용한 인사이트를 얻거나 글쓴이의 의도 또는 감성을 추출하기 위한 일련의 과정이다. 텍스트마이닝 기반의 분석을 추구한 이유는 우선 본 연구의 분석 대상인 “디지털 플랫폼 상 사용자들이 작성하는 리뷰 텍스트 및 평가”를 대량으로 얻을 수 있어 분석 결과의 신뢰성을 더욱 높일 수 있을 뿐만 아니라, 효과적이고 빠르게 분석하여 보다 더 빠른 피드백을 사용자에게 제공할 수 있다는 장점이 있기 때문이다. 이를 기초로 감성분석(Sentiment Analysis) 기법을 사용하여 사용자가 생성한 텍스트(User Generated Data) 속에 내포된 전기차에 대한

사용자의 감성을 식별하고자 하였는데, 더욱 세분화된 사용자 감성을 추출하고자 Aspect-Based Sentiment Analysis(ABSA) 방법을 적용하였다.

정리하자면, ABSA 방법론을 활용하여 수집된 전기차 사용자 리뷰 데이터로부터 전기차의 주요 구성 요소(Component)와 사용자와 전기차 간 인터랙션(User-EV Interaction)에 기초한 주요 특성(Attribute)을 추출하는 주요 요소 추출(Asspect Extraction)과 요소 별 사용자 감성 분석(Sentiment Analysis)을 주요 목표로 하며, 이 과정에서 수반되는 사용자 리뷰 부족 문제와 데이터 불균형(Data Imbalance) 문제를 극복하는 것을 하위 목표로 한다. 더불어 사용자 긍정/부정 감성에 영향력 있는 차량 요소(스펙)가 무엇인지 확인할 수 있는 새로운 UX 분석 프레임워크를 제안하고자 한다. Figure 2.1은 본 연구의 목표와 최종 산출물을 도식화한 것이다. 이로써 사용자 관점에서 주요 전기차 구성 요소와 사용 특성, 그리고 사용자 긍정/부정 감성에 영향력 있는 전기차의 스펙을 확인함으로써, 기존의 인터뷰나 설문조사에 얻는 인사이트 수준의 아웃풋을 얻고자 한다.

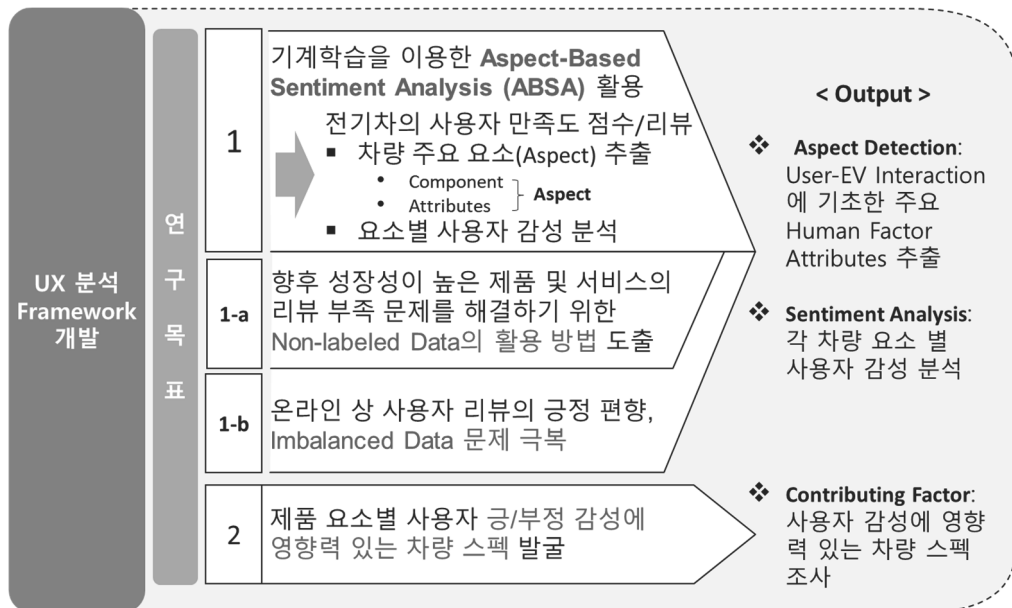


Figure 2.1 연구 주요 목표(1,2)와 하위 목표(1-a, 1-b) 및 최종 산출물

2.2 선행 연구

2.2.1 Aspect-Based Sentiment Analysis

ABSA(Asspect-Based Sentiment Analysis)란 대상의 하위 요소 및 특성과 관련된 용어를 고려하여, 각각에 대한 감정을 식별하는 기술이다. ABSA를 활용한 초기 연구는 “주요 어휘가 자주 반복된다”고 전제하면서 명사 출현 빈도로 Aspect를 추출하고 이에 대한 감성 극성을 계산하는 방법이었다 [4]. 한편 최근 가장 주목받고 있는 ABSA 적용 방법은 3가지 Subtask를 통한 문제 해결 방법으로, 이는 Aspect 추출(Aspect Extraction), Aspect의 카테고리 탐지(Aspect Category Detection), 그리고 Aspect에 대한 감성 극성(Sentiment Polarity)을 추출하는 것이다 [5]. 이 때 각 Aspect와 Aspect에 대한 감정을 추출하기 위해, 분석 텍스트로부터 얻은 전체 말뭉치(Corpus)에서 Aspect 카테고리화 및 이에 해당하는 Aspect 용어, 그리고 정답 감성 Label이 필요하다. 전형적인 감성 분석은 주어진 문장 또는 문장으로 이루어진 단락(Paragraph)에 대한 감성을 긍정/부정으로 분류 또는 예측한다. 대개 주어진 텍스트에 하나의 측면과 하나의 극성만 있다는 가정으로 진행되는 반면, 더 일반적이고 복잡한 Task는 문장에서 언급된 양상과 각각의 Aspect에 관련된 감성을 예측하는 것인데, 이를 Aspect-Based Sentiment Analysis (ABSA)라 한다 [4].

ABSA를 활용한 주요 과제는 문서 내 주요 Aspect를 발굴하고, 세분화된 감성을 추출하는 것을 일컫는다. 주로 각 문장 단위 또는 문서 단위로 Aspect 용어나 Aspect Category 레이블과 5점 척도의 사용자 감성 레이블이 함께 있는 데이터셋을 대상으로 하며, 여기에 연구자가 설계한 기계학습 모델을 기반으로 Aspect와 이에 대한 감성 극성(Sentiment Polarity)에 대해 예측 정확도를 높이고자 하는 Multi-Classification 문제를 해결하고자 하는 것이다.

데이터셋(Dataset)은 대부분 제품에 대한 웹 사이트의 사용자 리뷰이다. 초기 ABSA를 활용한 연구는 특정 제품에 대한 리뷰 데이터셋을 사용하는 경향이 있었으나, 최근에는 제품 리뷰 외에도 ABSA를 위한 차별화된 데이터 셋인

레스토랑, 영화, 호텔과 같은 서비스에 대한 리뷰도 사용되고 있다 [5, 6, 7, 8].

ABSA를 활용하는 과제는 매년 활발하게 진행되고 있는데, 대표적으로 SemEval(International Workshop on Semantic Evaluation)에서 2014년부터 Shared Task로써 참가자들 간 Contest 형식의 ABSA Workshop이 개최되고 있다. 여기서 Aspect와 Sentiment 각각의 레이블이 있는 여러 도메인의 ABSA Dataset에 대해 각 참가자 그룹의 방법론 기반으로 예측 정확도가 높은 순으로 평가한다. 그러나 실제 산업에서 분석하고자 할 때, Aspect 레이블이 있는 데이터는 거의 없다. 이처럼 Aspect 용어 레이블이 없을 경우에는, 비지도적 방법(Unsupervised Method)을 통해 Aspect를 추출하고 이에 대한 감성 분석을 하는 방법으로 접근한다 [9].

ABSA를 통해 추출한 정보는 일반적인 사용자 평가 또는 감성 분석을 적용하여 추출할 수 없기 때문에 상당히 가치 있는 정보이다. 일반적인 사용자 평가 또는 감성 분석은 제품에 대한 사용자의 일반적인 관점이나 감성을 제공하는데 ABSA는 특히 이 감정이 어떤 측면을 가리키는 지 파악하는 것을 목표로 함으로써 기존 방법의 결점을 보완할 수 있다 [4]. 대부분의 문단 또는 문서 단위 감성 분류는 일종의 사용자 감성의 Average out 효과를 나타나게 되는데, ABSA는 이처럼 일반적으로 Document 단위로 하나의 감정 분석 결과에서 제공함으로써 불러일으킬 수 있는 오해를 최소화하고, 사용자 경험을 훨씬 세분화되고 정확하게 분석하여 각 Aspect에 대한 사용자 경험 및 세부 피드백을 대량의 비정형 텍스트로부터 얻을 수 있다.

전통적인 감성 분석(Sentiment Analysis)은 문서 단위(Document Level)로 이루어졌으며, 각 제품 리뷰의 감성 극성(Sentiment Polarity)을 계산하는 것을 예로 들 수 있다. 이처럼 리뷰 텍스트 전체의 감성 극성 추출에 중점을 두었으나 점차 세분화된 감성 추출 방법이 요구됨에 따라, 연구자들은 주어진 제품 리뷰 텍스트에서 주제(Topic)를 발굴하거나, 제품의 특성(Attribute) 또는 주요 부품 요소(Part)를 추출하기 시작하였으며 [4,10], 이러한 주요 요소들에 대한 사용자 감성을 파악하기 위해, 단락(Paragraph) 또는 문장(Sentence) 단위 감성 분석을 통해,

해당 문서 및 문장이 어떠한 감정을 나타내고 있는지 파악하였다. 이러한 세분화된 접근 방식, 즉 Aspect-Based Sentiment Analysis (ABSA)는 감성 분석에서 중요한 방법론이 되었다 [6].

주제 단위 감성 분석은 ABSA 외에도 다양한 접근 방법이 있는데 이중 LARA [11]가 대표적이다. LARA는 전체 리뷰의 평가 점수(User's Rating)를 바탕으로 리뷰 내 여러가지 Aspect에 대해 평가 점수를 산출하는 방법으로, 전체 리뷰 평점과 Aspects간의 Weight를 Linear Regression 기반으로 산정하여 각 Aspects에 대한 평점을 도출하였다. 반면 ABSA는 전체 Rating과는 무관하게 각 주제의 문장이나 단락의 감성 극성을 Lexicon 또는 기계학습 기반으로 독립적으로 계산하는 방법으로 접근한다는 점에서 차이가 있다.

ABSA를 활용한 분석 과정은 연구자 뿐 만 아니라 잠재 고객과 제조 및 서비스 업체 모두가 체계적으로 관련 피드백에 액세스 할 수 있도록, 사용자 리뷰 마이닝 및 요약에 위한 워크 플로우가 존재한다 [12].

ABSA는 크게 3가지 중요한 Subtask로 나뉘는데 이는 Pontiki et al. [5]에서 처음 제안한 방법이다. 첫 번째는 Aspect Term을 추출하는 것이고, 두 번째는 주어진 Aspect Term들을 서로 다른 카테고리로 묶는 것이다. 세 번째는 Aspect Term 별 감성 극성을 추출하는 것인데, Aspect Term은 대개 한 문장 내에 위치하기 때문에 해당 문장 내 Aspect Term에 대한 감성 극성을 판정하는 방법을 취한다. 초기에는 감성 사전 기반의 Lexicon-based methods [6, 12]을 취하였으나, 기계학습 분야의 발전에 따라 점차 기계학습 기반의 감성 분석을 진행하고 있다 [13].

2.2.2 요소 추출(Asspect Extraction)

요소(Asspect)는 크게 구성 부품(Component)와 특성(Attribute)로 정의할 수 있다. 구성 부품(Component)이란 제품 또는 서비스의 구성요소 및 부품을 의미하는데 전기차로 예를 들면 배터리, 소프트웨어, 사운드 시스템(Sound System) 등이다. 특성(Attribute)이란 사용자가 느끼는 차량의 특성을 뜻하고, 전기차로 예를 들면

가속감(Acceleration), 안전성(Safety), 사용 편의성(Ease of Use) 등이다.

초기 연구에서의 요소 추출(Aspect Extraction) 접근 방법은 주요 토픽은 반복해서 등장한다는 전제하에, 출현 빈도의 명사 또는 명사구를 추출하는 방식이었다 [4]. 그 외에도 Rule-based Linguistic Pattern 기반으로 추출하는 연구가 있었는데 이는 문법적인 관계를 활용하여 문장의 구조를 파악하고, 구문적 특성을 활용하여 감성 어휘와 연관된 Aspect를 추출하는 방법이다 [14, 15]. 또한 LDA(Latent Dirichlet Allocation) 기반의 토픽 모델링(Topic Modelling)을 통한 토픽 추출 방법도 널리 활용되었다 [16, 17, 18]. 최근에는 딥러닝 기반으로 Aspect를 추출하는 시도가 있는데, 다만 지도 학습 기반으로 Label이 있는 데이터 셋에 한하여 적용되는 경우가 많다 [19, 20, 21].

지금까지 설명한 방법 외에, 사용자가 작성한 리뷰 텍스트로부터 주요 구성 부품과 특성을 텍스트 요약(Text Summarization) 기법을 통해 주요 키워드(Keyword) 또는 주제를 추출하는 방법으로 진행할 수 있는데, 이러한 텍스트 요약 기법은 크게 두 가지 접근 방법이 있다. 첫 번째, 추상적 접근(Abstractive Approaches)은 사람이 글을 읽고 짧게 요약하는 것처럼, 문서 집합 혹은 한 문서의 내용을 기반으로 요약문을 생성하는 방법이다. 주로 기계학습 기반으로 진행하며 대표적으로 Seq2Seq(Sequence to Sequence) 학습 모델이 있다 [22]. 최근에는 어텐션 메커니즘(Attention Mechanism)과 함께 활용된 모델들을 통해 많은 진전이 있는 분야이다. 다만 Seq2Seq 모델은 기본적으로 지도 학습(Supervised Learning)이기 때문에 인공 신경망으로 훈련하기 위해서는 텍스트 원문과 함께 각 리뷰 텍스트 별 요약 문장 또는 정답 키워드 레이블(Label)이 있어야 가능하다. 하지만 본 연구에서 다루는 데이터는 텍스트 원문을 요약한 정보가 없기에 본 추상화 접근 방법으로 접근할 수 없었다. 두 번째, 추출적 접근(Extractive Approaches)은 주어진 문서 집합 내에서 이를 대표할 수 있는 단어들이나 문장들을 선택하는 방법으로, 주어진 텍스트 내에서 실제 존재하는 단어와 문장 중 가장 정보력이 높은 어구(Informative Phrase) 또는 문장을 선택한다. 따라서 실제 원문의 내용 및 맥락과

동떨어진 요약 결과를 낼 가능성은 낮지만, 원문 내 표현으로 제한된다는 한계가 있다. 대표적인 추출적 접근 방법으로는 TextRank [23]가 있다. TextRank는 텍스트에 Graph-based Ranking 알고리즘 기반의 PageRank를 활용한 것이다. PageRank는 웹 상에 존재하는 하이퍼링크를 가진 웹 문서의 상대적 중요도에 따라 문서의 가중치를 부여하는 방법으로, 서로 간의 인용과 참조로 연결된 임의의 묶음에 적용해볼 수 있다. 즉, PageRank가 높은 웹페이지는 다른 웹 사이트로부터 링크를 받거나 다른 사이트가 많이 참조한 것으로 이해할 수 있다. 이러한 PageRank를 텍스트 문서 내에 적용하여 중요한 단어, 주요 어구(Key Phrase), 또는 문장의 중요도에 따른 가중치를 부여하는 한 것이 TextRank이다. 핵심 단어 선택을 위한 점수 산정을 위해, TextRank는 두 단어 간의 유사도를 정의하기 위해서는 두 단어의 동시 출현 빈도(Co-occurrence)를 계산하는데, 여기서 Co-occurrence는 문장 내에서 두 단어의 간격이 설정 값(window)인 횟수이다. TextRank를 통해 단어, 어구 또는 문장의 순위를 계산하는 식은 다음과 같다.

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j)$$

WS(Vi): 문장 또는 단어(Vi)에 대한 TextRank 값

wij: 문장 또는 단어 i 와 j 사이의 가중치

d: damping factor

위와 같이 WS(Vi)를 계산한 뒤 높은 순으로 정렬한다. d는 damping factor로, PageRank에서 웹 서핑을 하는 사람이 해당 페이지를 만족하지 못하고 다른 페이지로 이동하는 확률로써, TextRank에서도 그 값을 그대로 사용(0.85로 설정)하였다. 이와 같이 각 그래프에 PageRank를 학습하여 각 Vertex (단어 혹은 문장)의 랭킹을 계산하는 과정으로, 계산 후 높은 순으로 정렬하여, 랭킹이 가장 높은 단어 또는 문장이 키워드와 핵심 문장이 되는 것이다.

2.2.3 감성 분석(Sentiment Analysis)

텍스트 감성 분석은 컴퓨터 과학에서 활발하게 연구되고 있는 한 분야로써 텍스트로부터 사용자의 의견을 추출하여 규명하고, 긍정/부정 감성을 분류하는 것을 주요 목표로 한다. 이를 위해서는 텍스트로부터 단순히 감성 극성(Polarity)을 계산하는 문제가 아닌 NLP Task의 문제로써, 보다 Human-like Performance를 위한 통합적인 접근 방법이 필요하다 [24].

감성 분석(Sentiment Analysis)를 위한 접근 방법은 크게 3가지 카테고리로 나뉘는데, 첫 번째는 지식(사전) 기반 감성 분석(Knowledge(Lexicon)-based Approaches), 두 번째는 기계학습 기반 감성 분석(Statistical & Machine Learning based Approaches), 그리고 Hybrid Approaches이다 [25]. 지식(사전) 기반 감성 분석은 감성 어휘 또는 문법 기반의 분류로, Happy, Sad, Boring과 같은 분명한 감성 어휘들의 존재에 따라 텍스트를 분류하거나 언어학적인 구조나 어휘간 문법적인 관계성을 통해 분류하는 방법이다. 통계&기계학습 기반 감성 분석은 Naïve Bayes 나 Support Vector Machines과 같은 전통적인 기계학습 모델 및 Deep Learning 기법을 사용하여 텍스트의 감성을 분류하는 방법이다. 마지막으로 Hybrid Approaches는 위의 기계학습 기반의 방법과 지식 기반 접근 방법을 모두 활용하는 방법이다.

ABSA에서 감성 분석은 주로 텍스트 특성들, 예를 들어 bag-of-words, linguistic features 등을 추출하여 이를 전통적인 기계학습 분류기를 통해 학습시키는 모델링에 중점을 둔다. 다만 최근의 신경망 기반 딥러닝 모델의 빠른 발전으로 ABSA Task에도 딥러닝 기반의 모델이 적용되기 시작했으며 [13], CNN, LSTM 기반의 다양한 모델이 적용되었다. 특히 최근에는 Neural Attention Mechanism을 활용하여 감성 예측을 위한 Representation에 대한 Aspect의 영향력을 강화하고자 하였다 [26, 27]. ABSA Task에 적용된 Attention 기반 모델로는 IAN [28], MemNet [29], BILSTM-ATT-G [30], RAM [31]이 있었으며, 이후 MGAN(Multi-Grain Attention Network) [32]를 활용한 ABSA를 통해 가장 좋은 성능을 냈다.

2.2.3.1 기계학습 기반 감성 분석

기계학습 기반 감성 분석(Statistical & Machine Learning Based Sentiment Analysis)이란 기계학습 기반 알고리즘으로 텍스트에서 추출한 특징을 학습시킨 모델을 바탕으로 문장, 문단, 문서 단위의 감성 분류/예측하는 것이다. ABSA Task에서 주로 사용되는 기계학습 모델의 상세 구조 및 개념은 다음과 같다.

첫 번째는 CNN(Convolutional Neural Network)으로 NLP Task에서 자주 활용되는 딥러닝 모델이다. 이미지 또는 텍스트를 1차원 텐서인 벡터로 변환하고 다층 퍼셉트론(Multi-Layer Perceptron)의 입력층으로 사용할 시, 변환 전에 가지고 있던 공간적인 구조(Spatial Structure) 정보가 유실될 가능성이 있다. 여기서 공간적인 구조 정보라는 것은 텍스트로 예를 들면, 연속적인 단어로 이루어진 문장, 문서 내에서 단어 간 위치, 계층적인 구조와 같은 정보를 의미한다. 따라서 결국 이미지 또는 텍스트의 공간적인 구조 정보를 보존하면서 학습할 수 있는 방법이 필요한데, CNN이 해당 역할이 가능한 모델이다 [33].

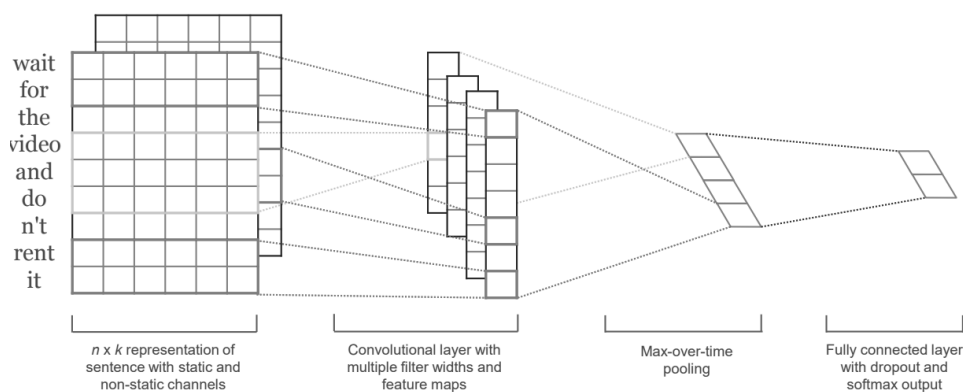


Figure 2.2 CNN for Sentence Classification [33]

CNN의 가장 큰 장점은 텍스트 내에서 중요한 N-gram 특징을 추출하여, 정보력 있는 의미 벡터(Informative Latent Semantic Representation Vector)를 생성하여 이를 분류 문제(Classification Task)에 활용할 수 있다. 1 layer의 기본적인 컨셉을 소개한

Yoon Kim [33] 이후 ABSA Task에도 CNN을 적용한 많은 연구들이 있다 [19, 34, 35].

두 번째는 어텐션 기반 양방향 장단기 메모리 네트워크(**Bidirectional-Long Shot Term Memory with Attention**)이다. LSTM은 기존의 RNN에서 발생하는 데이터 길이가 길고 층이 깊을 경우 과거의 정보가 손실되는 **Vanishing Gradient** 문제를 cell state 도입을 통해 극복한 알고리즘이다. 한편 RNN이나 LSTM은 입력 순서를 시간 순대로 입력하기 때문에 결과물이 직전 패턴을 기반으로 수렴하는 경향을 보인다는 한계가 있다. 이 단점을 해결하는 목적으로 양방향 순환 신경망(**Bi-RNN**)이 제안되었으며, Bi-RNN은 기존의 순방향에 역방향을 은닉층에 추가하여 성능을 향상시켰다. 어텐션 기반 양방향 장단기 메모리 네트워크를 Classification Task에 적용한 이후[36], 이를 ABSA에 활용한 후속 연구들이 등장하였으며 좋은 성능을 이끌어냈다 [37, 38].

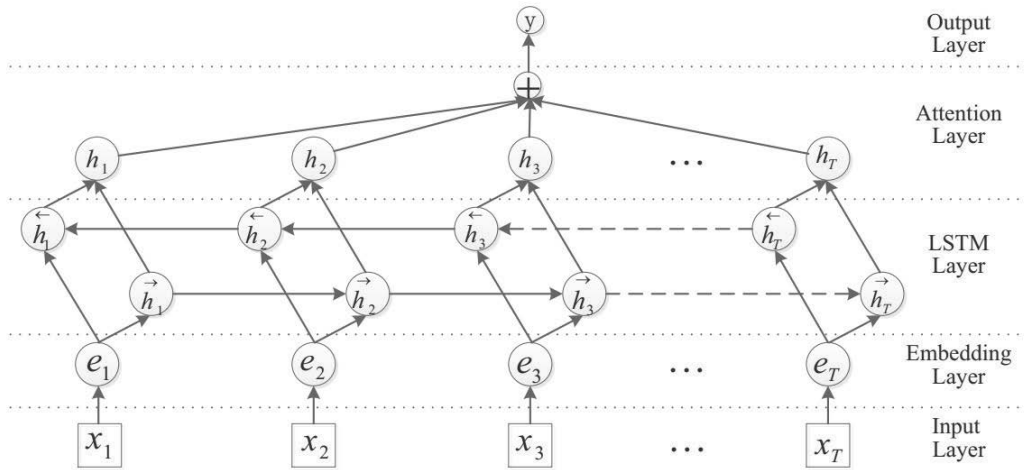


Figure 2.3 Attention-based bidirectional LSTM for classification [36]

세 번째는 **Hierarchical Attention Network(HAN)**이다. Attention Network는 디코더(Decoder)에서 출력 단어를 예측하는 매 시점(Time Step)마다, 인코더에서의 전체 입력 문장을 다시 한 번 참고한다는 점이 특징이다. 단, 전체 입력 문장을 전부 다 동일한 비율로 참고하는 것이 아니라, 해당 시점에서 예측할 단어와

연관이 있는 입력 단어 부분을 좀 더 집중해서 본다는 것이 Attention의 주요 아이디어이다. 특히 본 모델은 Word, Sentence 각각 Attention Layer를 추가함으로써 언어적 구조를 잘 반영한 모델이라고 볼 수 있다.

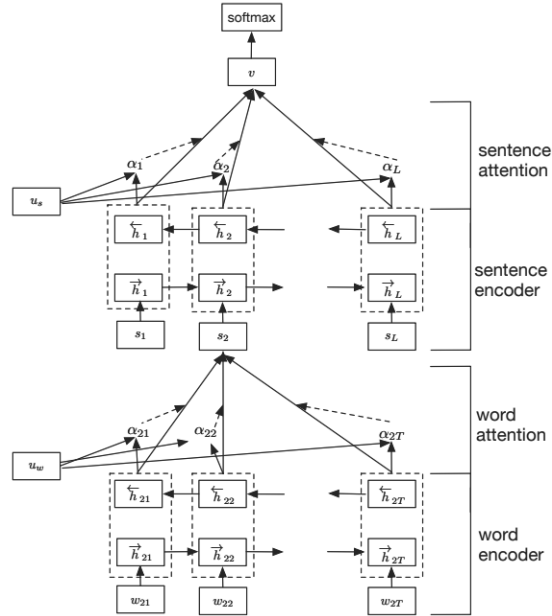


Figure 2.4 Hierarchical Attention Network for Text Classification [39]

기본적으로 하나의 문서는 단어 -> 문장 -> 문서 순으로 이어지는 Hierarchical structure를 갖고 있기 때문에 이러한 특징을 학습할 수 있는 네트워크가 문서를 학습하고 분류하는 데 적합한 네트워크라고 볼 수 있는데, HAN [39]은 이에 아주 부합하는 모델로 평가되고 있으며, 이를 ABSA Task에 활용한 연구에서도 좋은 결과를 보여주었다 [40].

네 번째는 준지도학습(Semi-Supervised Learning)이다. 준지도학습은 텍스트로부터 추출된 특징들을 학습하여 감성을 예측하는 직접적인 알고리즘은 아니지만 앞서 언급한 모든 알고리즘에 Wrapper Method로써 적용이 가능하며,

성능을 개선시킬 수 있는 중요한 알고리즘이다. 머신 러닝(machine learning) 방법론들 중 지도 학습(supervised learning) 모델들은 정답(Label)가 꼭 있어야만 학습이 가능하다는 단점이 있는데, 이를 극복하기 위해 제안된 방법이 준지도 학습(semi-supervised learning)이다. 이는 Figure 2.6의 설명과 같이, 정답 Label이 있는 데이터로 학습된 분류 모델(Teacher Model)을 활용하여, Non-label 데이터를 Teacher Model에 입력 값으로 넣으면서 예측 값을 구하는데, 이 중 높은 확률 값이 나오는 데이터 위주로 Pseudo Label을 부여하는 방법으로 Label 전파(Propagation)를 진행한다. 이처럼 새롭게 부여된 Label의 데이터와 기존의 정답 Label이 있는 데이터를 모두 사용하여 새로운 모델(Student Model)을 재학습(Re-training) 시키면서 분류 성능을 향상시키는 방법을 셀프 트레이닝(Self-Training)이라고 하며, 이는 Semi-Supervised Learning의 가장 간단한 모델이다. 와 같은 준지도학습(Semi Supervised Learning) 방법은 모든 알고리즘에 Wrapper Method로써 적용이 가능하다.

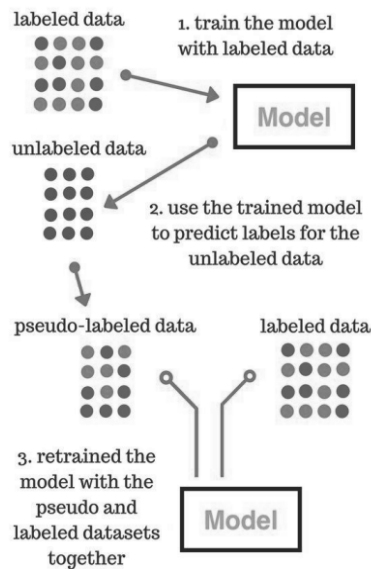


Figure 2.5 Self-Training [41]

2.2.3.2 Lexicon 기반 접근 방법

사전에 정의된 긍정/부정 단어를 이용하여, 텍스트에 포함된 단어의 감성 극성(Sentiment Polarity)을 일정한 규칙 기반으로 텍스트의 긍정과 부정을 판별하는 일종의 Rule-based 방법이다. 즉, 문장 내에 있는 단어들의 감성 사전 내 점수 기반으로 전체 문장의 감성 점수를 계산하는 방법으로, 사용자의 긍정/부정 Label이 없어도 감성 어휘 사전(Dictionary)만 있다면 문장 또는 문단의 긍정/부정 판별이 가능하다. Label이 없는 문장, 문서에도 쉽게 활용될 수 있어 기계학습 기반의 감성 분류 이전에는 널리 활용되었다.

제 3 장 연구 방법

Aspect-Based Sentiment Analysis라는 세분화된 기술을 사용하여 전기차 사용자 리뷰 내에서 주요 Aspect 선정 및 해당 Aspect과 관련된 의견 식별 및 감성 극성 추출하고자 한다. 아래 Figure 3.1와 같이, 주요 Aspect가 A라고 가정하면, A에 포함된 모든 문장들(1, 3)을 불러와서 ML 모델 또는 Lexicon 기반으로 감성 극성을 판별한다. 그리고 주요 Aspect의 상세 스펙과 Aspect 대한 사용자의 감성 간의 상관관계를 확인해보고, 회귀분석을 통해 영향력 있는 변수가 무엇인지 탐색한다. 예를 들어, 주요 Aspect가 차량의 Size라면, 차량 Size에 관련된 주요 변수(Wheel Base의 크기, 옆면 길이, 후면 길이, 차량 높이 등)들을 Crawling해서 변수와 차량 Size에 대한 사용자의 긍정/부정 감성 간의 관계를 확인해보는 것이다.

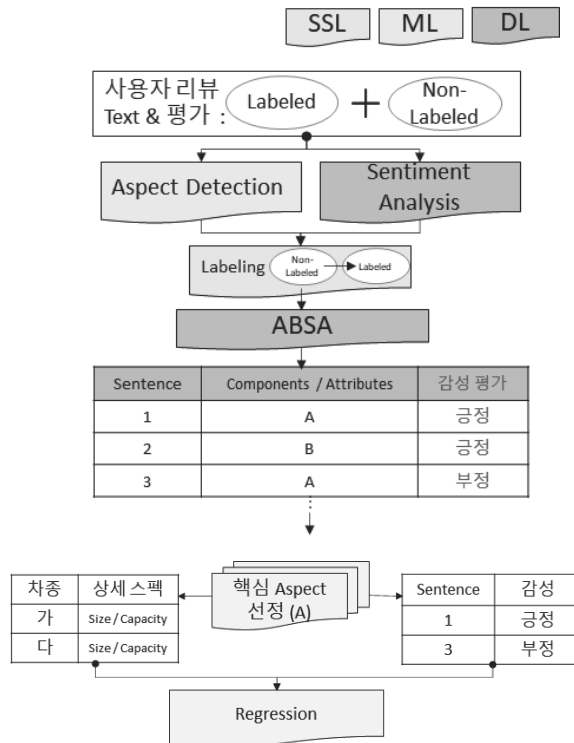


Figure 3.1 ABSA 기반 사용자 경험 분석 Framework

3.1 데이터 수집

수집 데이터의 종류는 크게 정답 Label이 있는 Label Data와 정답 Label이 없는 Non-Label Data 2가지로 분류되는데, 이는 각각 자동차 포럼 상의 리뷰 별로 사용자 만족도 Label이 존재하는 리뷰 텍스트 Dataset과 Youtube.com의 사용자 만족도 Label이 없는 Comment Dataset이다.

3.1.1 Label 데이터 수집

Label 데이터란 사용자의 만족도 Label이 존재하는 리뷰를 의미한다. 본 연구에서는 자동차 포럼(Car Forums)에서 수집하였는데, 유명 자동차 포럼에 모두 접속하여 가능한 많은 사용자 리뷰를 보유하고 있는 Website를 선별하였다. 그 결과, 사용자 리뷰 수가 가장 많은 Edmunds.com과 Cars.com을 중심으로 Cargurus.com, Carfax.com, Carbuyers.co.uk 까지 총 5개의 자동차 포럼에서 Data를 수집하였다. 위 자동차 포럼 내에서 현재 확인 가능한 모든 전기차종을 대상 차종으로 하였다. 데이터 수집은 Beautiful Soup이라는 파이썬 라이브러리를 활용하였는데, 각 자동차 포럼 내 주요 정보를 추출하고 Query 시 효과적으로 정보를 추출하기 위한 양식의 데이터 프레임을 별도 구축하여 이에 저장하였다.

가능한 많은 데이터를 얻기 위한 노력의 결과, 152개 차종에 대하여 총 5,065 건의 Label이 있는 사용자 리뷰를 수집하였다.

3.1.2 Non-Label 데이터 수집

Non-Label 데이터란 사용자의 만족도 Label이 존재하지 않는 리뷰로, 나름의 정형화된 양식으로 작성해야 하는 자동차 포럼 상 리뷰가 아닌 SNS나 웹에서 쉽게 접할 수 있는 텍스트 리뷰를 의미한다. 본 연구에서는 최근 가장 영향력 있는 소셜 네트워크로 꼽히는 Youtube를 대상으로 하였다. 전문 차량 Reviewer가

실제 전기차에 대한 상세 후기를 영상으로 업로드하면, 이를 시청한 누리꾼들이 남긴 코멘트(Comment)를 수집하였다. 대상 차종은 자동차 포럼에서 얻은 현존하는 차량 List를 동일하게 적용하였다. 데이터 수집은 Selenium API를 활용하여 자동 검색 및 Scroll Down하여 사용자 리뷰를 Crawling 하는 방법으로 진행하였다. 우선 데이터를 얻고자 하는 대상의 전기차와 Plugged-in-Hybrid 차량의 "제조사 - 모델명 - 생산 년도"를 모두 저장하고, Youtube.com 검색 키워드를 저장한 정보 내의 "제조사명 + 모델명 + 생산 년도 + Review" 자동 조합(예를 들면 "Tesla + model-s + 2018 + review")으로 한다. 해당 검색어가 들어간 Car Reviewer 스트리밍의 User Comment들을 수집했다. 저장 Data Frame은 Label Data와 동일하며, Car Forum에서는 얻을 수 있으나, Youtube.com에서 얻을 수 없는 정보들(i.e. 전문가 리뷰, 차량 상세 스펙)은 공란(Null)으로 두었다. 다만 Youtube.com으로부터 얻는 Non-labeled 사용자 의견들을 모두 사용할 수는 없고, 본 연구 목적에 맞게 UX 관련 내용의 Comment만 선별할 필요가 있기 때문에 UX Keyword가 있는 Comment만 선별했다. UX Keyword는 "I bought", "I ordered", "I buy", "my experience", "I experienced", "I choose", "I chose", "ve chosen", "driven", "I drove", "I purchased", "I own", "ve own", "I am owner", "I use", "ve used", "I used", "I was satisfied", "ve been satisfied"이다. 전기차 Model 명이 반드시 들어가는 리뷰만 선별하여 Comment에서 사용자 경험이 해당 스트리밍 제목의 차종의 Model일 가능성을 높였다.

획득한 Non-Label 데이터 중 유효한 데이터 수는 218,001개로, 가운데 Filtering을 거쳐 총 6,488개를 선별하였다.

3.1.3 데이터 분포

아래 Figure 3.2는 획득한 데이터의 분포를 나타낸 그래프이다. Figure 3.2의 좌측 푸른색 그래프가 Label 데이터의 분포이고, 우측 붉은색 분포 그래프가 Non-Label 데이터의 분포이다.

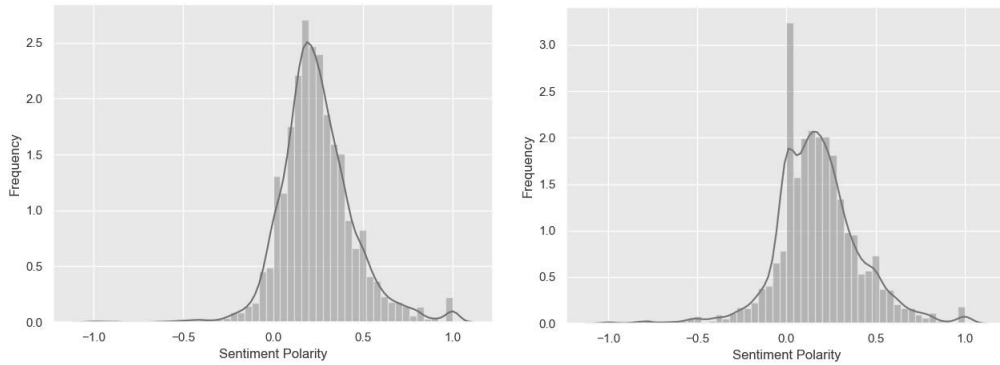


Figure 3.2 수집된 데이터의 Lexicon 기반 감성 분포
(좌 Label Data | 우: Non Label)

Lexicon 기반으로 감성 분포를 확인해본 결과, Label 데이터의 경우 (Figure 3.2 좌측) 총 5,065개의 유효 데이터 중에서 부정 323건, 중립 96건, 긍정 4,637건으로, 중립과 부정의견의 비율이 8.4%(총 428건)에 불과했다. 반면, Non-Label 데이터의 경우(Figure 3.2 우측) 총 6,488개의 유효 데이터 중 부정 881건, 중립 566건 그리고 긍정 5,041건으로, 중립과 부정의견의 비율이 22.4%(1,447건)이었다. 이는 SNS 상 자유롭게 게재하는 의견이 자동차 포럼상의 사용자 경험보다 덜 편향된 데이터임을 보여주는 사례로, SNS 상 사용자 의견을 함께 활용함으로써 긍정으로 치우친 제품 리뷰의 편향성을 다소 완화할 수 있을 것으로 보인다. 단, Youtube Comment 절대적인 수치로는 긍정 편향되어 있기에, 데이터 불균형 문제를 반드시 해결해야 보다 더 나은 분류모델을 구축할 수 있다.

3.2 데이터 전처리

수집된 텍스트들을 대상으로 삭제(Remove) 및 교체(Replace) 작업을 진행했다. 이는 기본적으로 SNS상의 텍스트들이 갖고 있는 양식 및 고유 특성들을 고려하여, 이를 텍스트 분석에 적합하도록 일부를 삭제(Remove) 또는 교체(Replace)하는 과정이며, 본 연구에서 적용한 기술들은 다음과 같다.

Remove

- ✓ URLs("https://", User Mentions(@) and Hashtags(#))
- ✓ Unicode Strings and Noise
- ✓ Numbers
- ✓ Emoticons
- ✓ Punctuation
- ✓ Parenthesis, Apostrophe, Quotation Mark
- ✓ Stopwords: # 불용어 사전 Source = <https://www.ranks.nl/stopwords>

Replace

- ✓ Replace Slang and Abbreviations: Twitter Slang 및 약어집 활용하였다.
- ✓ Replace Contraction: i.e. {"ain't": "is not", "aren't": "are not", "'cause": "because"}
- ✓ Replace Multi Exclamation Mark, Multi Question Mark, Multi Stop Mark
- ✓ Replace Repetitions of Punctuation: string.punctuation내 문자가 반복 "....."
- ✓ Replace Negations with Antonyms
- ✓ Replace Capitalized Words to Lowercase
- ✓ Replace Elongated Words: i.e. "Hellooooooooo~", "gooooooooood!!"
- ✓ Spelling Correction: Spello [42]을 활용하여, 철자를 교정하였다.

즉, 텍스트를 더욱 효율적으로 임베딩하여 학습 모델이 더 나은 성능이 도출될 수 있도록, 필요 없는 어휘와 불용어를 삭제하고, 어휘를 표준화하는 작업이다.

두 번째로, 텍스트를 단어 단위로 Split하는 토큰나이징(Tokenizing)을 거쳐, 품사 태깅(Part of Speech Tagging)을 Stanford Corenlp를 활용하여 진행하였다. 세 번째로, Token별 품사 정보를 바탕으로 표제어 추출(Lemmatization)을 진행하였다. 여기서 Lemmatization이란 표제어(Lemma)추출을 의미하며, Lemma는 '기본 사전형 단어' 또는 "형태소" 정도의 의미를 갖는다. 표제어 추출은 각 단어들의 표제어를 찾아가는 과정이다. 본 연구에서는 주로 명사와 형용사가 중요한 개체(Entity)이기에, 토큰화 된 텍스트는 lemmatization을 통해 주로 명사와 형용사로 일반화된다. 표제어 추출 및 토큰나이징은 Python의 nltk 라이브러리를 활용하였다.

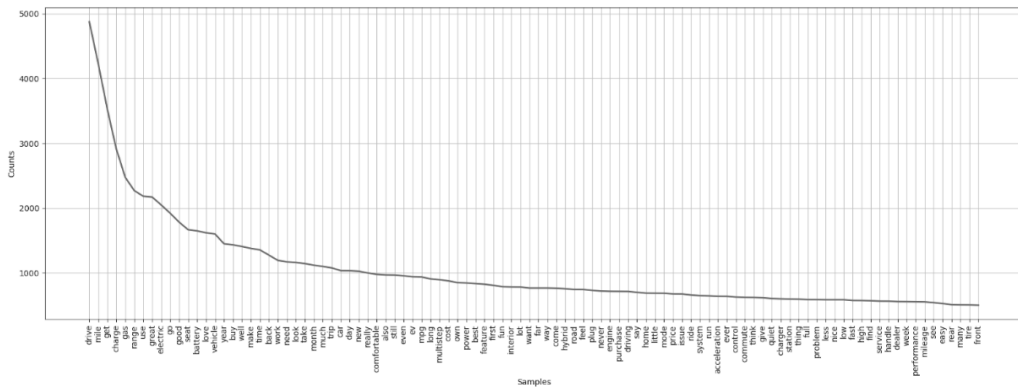
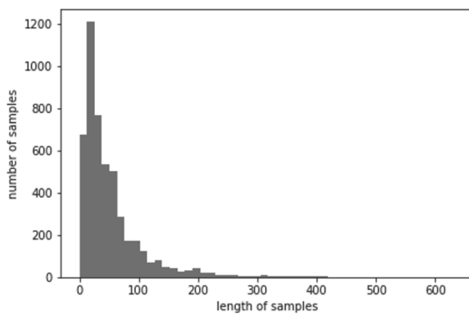


Figure 3.3 전처리 결과 I: Lemmatization 후 주요 단어

Figure 3.3은 전처리 후 추출한 표제어(Lemma)들을 빈도수 순서대로 나열한 그래프이다. 자동차 관련 리뷰만큼 “Car”, “Drive”, “Mile”, “Charge”와 같은 단어가 가장 많이 나왔다. 다만 “Car”를 비롯하여 쉽고 흔하게 등장하는 어휘나 텍스트 분석에 크게 중요하지 않다고 판단되는 단어는 불용어(Stopwords) 목록에 추가하였다. 그리고 본 연구는 자동차 모델 간 관련성 및 유사성을 파악하고자 하는 곳이 아니므로, 자동차 제조사명 그리고 모델명 또한 불용어(Stopwords) 목록에 포함시키는 것이 바람직하다. Labeled Data에 대해 전처리 전후 비교 결과는 Figure 3.4와 같으며, 전처리 전후로 고유 단어의 수가 23,351개에서 14,390개로 확연히 줄어든 점을 확인하였는데, 이는 상기 다양한 기술을 활용한 텍스트 전처리를 통해 단어 Indexing이 효과적으로 이루어졌음을 보여준다.



❖ 총 데이터 11,472개에 대한 전처리 결과

- Total Words : 1,173,406 Before preprocess
 - Total Distinct Tokens : 23,351 Before preprocess
- ↓
- Total Words : 515,175 After preprocess
 - Total Distinct Tokens : **14,390** After preprocess

Figure 3.4 전처리 결과 II: 문장의 길이 분포(좌)와 총 고유 단어의 개수(우)

- 총 문장의 개수: 78,732개
- 단어 집합(vocabulary)의 크기: 14,390
- 등장 빈도가 2번 이하인 희귀 단어의 수: 6,596
- 단어 집합에서 희귀 단어의 비율: 45.84%
- 전체 등장 빈도에서 희귀 단어 등장 빈도 비율: 2.4526459240775194

등장 빈도가 2회 미만, 즉 1회 밖에 되지 않는 단어들은 단어 집합에서 약 45.84%로 약 절반 정도의 비중을 차지함을 알 수 있다. 하지만, 실제로 훈련 데이터에서 등장 빈도로 차지하는 비중은 상대적으로 적은 수치인 2.45%밖에 되지 않는다. 등장 빈도가 지나치게 낮은 단어들은 자연어 처리에서 제외하고 싶다면, 단어 집합의 크기를 제한할 수 있지만, 본 연구에서는 별도로 단어 집합의 크기를 제한하진 않았다. 전체 단어 집합의 크기는 패딩을 위한 토큰인 0번 단어를 고려하며 +1을 해서 저장해주었다.

3.3 Aspect-Based Sentiment Analysis 수행

전처리가 완료된 Dataset에 대해 ABSA 방법론 기반 분석을 진행한다. ABSA는 크게 각 리뷰에서 내재된 모든 제품 Component와 Attribute를 찾는 Aspect Extraction 과정과 Sentence 단위 감성 예측 모델을 위한 Sentiment Analysis 과정 두 가지로 나뉜다.

ABSA 수행 시 고려해야할 점은 어느 수준(Level)에서 감성 분석을 진행할 지 결정하는 것이다. 텍스트는 언어학적으로 계층적 구조(Hierarchical Structure)로 이루어져 있기에, 어떤 수준(level)에서 분석할지 고려해야한다. 기본적으로 텍스트의 계층적 구조에 따라 문서 (Document), 문단 (Paragraph), 문장 (Sentence), 그리고 단어 (Word) 수준으로 분석할 수 있으며, 여기서 ABSA는 측면(Asspect) 단위로 사용자의 감성을 분류하는 방법이다. 다만, 데이터 셋에 따라 다르게 진행할 수 있는데, 예를 들어 Aspect(Component/Attribute) Label이 문단 단위로 있는

데이터 셋의 경우 문단 레벨의 감성 분석을 진행하며, 문장 단위로 있는 데이터 셋의 경우 문장 단위 감성 분석을 진행한다. 현재 다양한 주제 (예를 들어 레스토랑, 노트북, 휴대폰 등) ABSA 용도의 오픈 데이터셋을 다방면의 연구 목적으로 활용되고 있는데, 이러한 데이터 셋들은 각각 문장 또는 문서 단위로 Aspect 정답 Label이 존재한다.

3.3.1 요소 추출(Asspect Extraction)

우선 Aspect 추출을 위해 크게 두 가지의 추출적 접근 방법(Extractive Approach)을 적용하였는데, 첫 번째는 TextRank이며, 두 번째는 Naïve Method이다. 두 방법 모두 명사구(Noun Phrase)를 추출하기 위한 목적인데, TextRank는 N-grams 및 Collocation을 얻기 위한 목적으로 활용했고, Naïve Method는 존재하는 모든 독립 어휘(명사)를 추출하는 데 조금 더 주안점을 두었다.

3.3.1.1 TextRank

TextRank는 Word Graph (또는 Sentence Graph)를 구축한 뒤, Google 검색의 핵심인 Graph Ranking Algorithm 기반의 PageRank를 텍스트에 적용하여 문서 내의 키워드(또는 핵심 문장)를 Ranking 순으로 추출하는 방법이며, 본 모델을 구현하기 위한 상세는 다음과 같다.

첫 번째, TextRank는 그래프 기반 모델이므로 그래프를 만들어야 한다. 각 어휘는 그래프의 Vertex 역할을 하며, 전체 어휘 목록 내의 어휘 Index가 Vertex들 상에 표시된다. Weighted Edge Matrix는 모든 Vertex들 사이의 Edge 연결 정보를 모두 포함하고 있다. 본 방법에서는 Weighted Undirected Edges를 활용하였고, Weighted_Edge [i][j]는 어휘 인덱스 i로 표현된 단어 Vertex과 어휘 j로 표현된 단어 Vertex 사이의 연결 Edge의 Weight 값을 저장하고 있다. 즉, Weighted Edge [i][j]가 0이면 인덱스 i와 j로 표현되는 단어 사이에 Edge 연결이 없음을 의미한다.

텍스트에서 지정된 'Window_Size'의 Window 내에서 단어가 공존하는 경우, 단어 간에 (따라서 단어를 나타내는 i와 j간) 연결이 있다. $Weighted_edge[i][j]$ 내의 Connection Edge 값은 텍스트 내에서 서로 다른 위치에서 동일한 단어 사이에서 발견된 모든 연결에 대해 “ $1/(단어를\ 나타내는\ i와\ j간\ 의\ 거리)$ ” 만큼 증가한다.

그리고 Covered_Co-occurrences 목록(해당 위치에서 co-occurrence가 이미 확인된 단어의 문서 내에서의 pairs of absolute positions 목록)을 생성하여 관리함에 따라, 한 번에 하나의 텍스트 단위로 Sliding Window시에 텍스트 내의 동일한 위치에 있는 동일한 두 단어가 반복적으로 계산되지 않도록 하였다. 모든 Vertex 상의 Score는 1로 초기화 시키고, 단, Self-connections 고려되지 않으므로 $Weighted_Edge[i][i]$ 는 0이 된다. 그리고 Vertex와 연관된 모든 Undirected connection과 Edge들의 Weighted 합이 계산된다.

두 번째 Vertex의 TextRank 점수를 매기고 Key phrases의 순위를 확인하는 과정이다. Vertex i에 대해 연결되어 있는 Vertex j에 대한 Scoring은 앞서 2.2.2에서 언급한 TextRank 공식으로 계산한다. 따라서

$$Score[i] = (1 - d) + d[Sum(j) * \left(\frac{Weighted_edge[i][j]}{inout[j]} \right) * score[j]]$$

으로 정의할 수 있으며, 여기서 d는 damping factor이며 해당 Score가 수렴할 때까지 iteration을 돌면서 업데이트 된다. 마지막으로 Scoring Key phrases 후보군들에 대해 Scoring을 진행하고 내림 차순으로 정리한다. 가장 위에 있는 Keyword가 Key Phrase이다.

3.3.1.2 Naïve Method

두 번째 Naive Method는 형용사(Adject)와 명사(Noun)에 한하여 연속적으로 등장하는 모든 단어 조합을 수집한 후, Redundancy Pruning을 거쳐 명사구(Noun Phrase)와 연어(Collocation)를 얻는 방법이다. 본 방법은 기본적으로 자주 등장하는 명사 및 명사구를 찾는 방법과 유사한데, 이는 사용자들이 한 개체(Entity)의 서로

다른 측면에 대해 의견을 말할 때 일반적으로 사용되는 어휘가 수렴한다 [4]는 가정을 전제로 하였다.

연속적으로 등장하는 명사(명사구) 수집을 통해 전기차의 부품 용어나 사용자들이 자주 사용하는 전기차 관련 용어를 얻고자 했으며, 형용사를 통해 사용자가 전기차에 대해 느끼는 주요 감성 어휘를 얻을 수 있을 것으로 기대했다. 또한 형용사 + 명사 조합의 연어(Collocation)를 얻을 수 있기에 추출 어휘를 형용사와 명사로 제한하고 다른 품사는 모두 불용어(Stopwords)에 추가하였다. 이와 같은 방법은 별 다른 의미론적으로나 또는 단어 위치 기반으로 Count하는 것이 아니고 모든 명사(1개 이상 연속적으로 등장하는)를 고려하는 방법이기에 Naive Method라고 명명하였다. 이로써 추출된 명사구 및 연어에 대하여, 공통단어(Common Word)의 비중이 50%이상인 경우 Redundancy가 높다고 판단하여 제거하였고, 공통단어의 기준은 해당 단어가 나타나는 빈도가 전체의 10%를 넘을 경우로 정의하였다. 이러한 Redundancy Pruning 후, 일정 Threshold 이상 어휘만을 선별하였는데 본 과정에서는 5번 이상 등장할 경우로 한하였다. 본 알고리즘의 정확도를 높이기 위해 PMI 점수를 활용하는 방법 [43] 또한 있으나 본 연구에서는 적용하지 않았다.

3.3.2 감성 분석(Sentiment Analysis)

본 논문에서 다루는 데이터셋은 문단 또는 문장 별 Aspect 정답 Label이 존재하지 않는 데이터이기에, Aspect가 포함된 “문장(Sentence)” 레벨의 감성 극성(Sentiment Polarity)를 파악하였다. Aspect-level 감성 분석을 위해, 기계학습 모델 기반으로 Aspect가 포함된 문장 단위 긍정/부정 예측 모델링을 수행하고자 하였다. 단 Label Data가 충분하지 않아 문장의 감성이 정확하게 분류되지 않을 수 있는 점을 고려하여, 이를 Lexicon based 감성 분석을 병행하여 보완하고자 하였다.

3.3.2.1 기계학습 기반 감성 분석

본 연구에서는 Review 단위로 사용자 만족도 Label이 있는 텍스트를 학습시킨 기계학습 모델 기반으로 문장의 감성을 예측하였다. 중요한 점은, 리뷰의 첫 한 두 문장 (약 150자 내)의 텍스트와 리뷰 단위 평점을 학습 데이터로 사용하고, Aspect가 포함된 문장에 대한 Inference를 진행하는 방향으로 진행했다는 점이다. 이는 리뷰의 첫 한 두 문장만 봐도 문서 전체의 극성을 판단할 수 있다는 전제 하에 진행하였으며, 실제로 문장의 길이가 1041 이하로 문장의 감성을 판단할 수 있다고 한 문헌의 내용을 참조하였다 [44]. 이처럼 리뷰 단위로 사용자 만족도(감성) Label이 있는 데이터를 대상으로, 텍스트를 단어 기반의 임베딩을 통해 기계학습 모델에 학습시키는 과정을 거쳤으며, 상세 과정은 다음과 같다.

첫 번째는 임베딩을 위한 전처리 과정이다. 먼저 전처리가 완료된 리뷰 텍스트들의 평균 길이보다 약간 작게 문장의 길이를 제한하였다. 일종의 절단된 리뷰(Truncated Review)로써, 본 실험에서는 문장 길이를 약 150 단어로 제한하여 절단하였다. 단어 150개를 초과할 시 150개 이후의 텍스트는 모두 무시하고, 리뷰의 길이가 이보다 짧으면 Zero Padding을 통해 길이를 모두 동일하게 맞추었다.

두 번째는 문장 임베딩 과정이다. 단어를 밀집 벡터(Dense Vector)의 형태로 표현하는 방법을 워드 임베딩(Word Embedding)이라고 하며, Word2Vec, FastText, Glove, Bert 등 여러가지 방법이 있는데, 기존의 임베딩 방법의 단점을 극복해가는 방향으로 점차 발전해가고 있다. 본 연구에서는 총 세 가지 임베딩 방법을 사용하였고, 각 모델 별로 가장 효율적인 임베딩 조합이 무엇인지 모색해봤다. 첫 번째로 TF-IDF(Term Frequency-Inverse Document Frequency)는 말 그대로 단어의 빈도(Term Frequency)와 역 문서 빈도(Inverse Document Frequency)를 사용하여 문서 내의 각 단어들마다 중요한 정도를 가중치로 주는 방법이다. 단어의 의미를 밀집 벡터 형태는 아니지만, TF-IDF를 통해 문서 내 존재하는 단어 별 중요도가 결국 문서 분류에 중요한 영향력이 있다는 점과 긍정/부정 문서의 유사도 차이를 바탕으로 분류할 수 있다는 점에 있어서 이를 선택했다. 우선 희소표현(Sparse

Representation)인 문서 단어 행렬(DTM)으로 만들고 이를 TF-IDF 가중치를 부여하는 방법으로 사용할 수 있다. 두 번째로 GloVe [45]를 사용하였는데, 사전 학습된 어휘 벡터를 활용하여, 모든 단어가 100차원의 밀집 벡터 형태로 표현될 수 있도록 하였다. GloVe를 선택하게 된 이유는 다음과 같다. Word2Vec(Skip-Gram) 그리고 GloVe 모두 밀집 표현으로 임베딩 된 단어벡터 간의 유사도를 측정하기에 의미적인 유사성 및 위치 정보를 활용할 수 있다는 장점이 있지만, 사용자가 지정한 윈도우 내에서만 학습이 이뤄지기 때문에 말뭉치 전체의 동시 출현(co-occurrence) 정보는 다소 반영되기 어려운 단점을 지닌다. 다만, 본 연구는 텍스트 전체 말뭉치 내에서 동시 등장확률을 고려하고자 하는 컨셉으로써, 임베딩된 두 단어벡터의 내적이 말뭉치 전체에서의 동시 등장확률의 Log scale값이 되도록 목적함수를 정의한 GloVe를 선택하게 되었다. 또한 단어의 의미 기반으로 긍정/부정을 분류하고자 하는 만큼, Character Level의 N-gram은 본 논문에서 고려할 필요가 없었으며, User Generated Contents를 다루기에 Out of Vocabulary 문제는 크게 발생하지 않을 것으로 보아 FastText와 같은 임베딩 방법을 고려하진 않았다. 사전 학습된 어휘 표현은 Pre-trained GloVe Twitter 27b 100d를 참조하였다. 마지막 임베딩 방법은 Random Initialization으로 단어벡터의 초기값으로 무작위의 Index값을 부여하는 방법이며, Index 값은 사용자가 지정한 단어벡터 차원의 수 중 하나가 될 것이다. 단어의 중요도나 의미 유사성을 알기 어렵기에, Random하게 초기화하고 Embedding층에서 다른 파라미터들처럼 학습 과정을 통해 텍스트 분류를 가장 잘하는 방식으로 업데이트해서 사용하기 위함이다.

세 번째는 Imbalanced Data 극복을 위한 Data Augmentation 과정이다. 데이터 클래스 비율의 차이가 큰 경우를 데이터 불균형(Imbalanced Data)라고 하며, 이 경우 단순히 우세한 클래스 쪽으로 예측이 쏠리는 현상이 발생하여 모형의 정확도가 높을 수 있으나, 클래스 분류 성능이 좋아지지 않는 현상이 발생한다. 즉, 이러한 비대칭 데이터셋에서는 정확도(accuracy)가 높아도 재현율(recall) 및 F1 Score가 급격히 떨어지기에, 결국 성능이 좋지 않은 모델을 구축할 가능성이 높아

반드시 해결해야할 문제이다. 대표적인 해결 방법으로는 언더샘플링(Under sampling)과 오버샘플링(Over Sampling) 방법이 있다. 언더샘플링을 할 경우 Minority Class의 개수에 맞추기 위해 Majority Class에 있는 데이터의 일부를 샘플링하게 되는데, 이 경우 전체 데이터 사이즈가 1,000개 이하로 지나치게 작아지기 때문에 본 연구에서는 오버샘플링 방법을 선택하였다. 오버샘플링 방법 중, 무작위로 소수 데이터를 복제하는 무작위 추출이나 사전에 기준을 정해서 소수 데이터만 복제하는 방식이 있으나, 이는 과적합(Overfitting) 문제를 야기할 가능성이 높기에 결국 합성 데이터를 생성하는 방식을 선택하였다. 이중에서 SMOTE(Synthetic Minority Oversampling TEchnique) [46] 알고리즘은 데이터의 개수가 적은 클래스의 표본을 가져온 뒤 임의의 값을 추가하여 새로운 합성데이터를 생성하여 데이터에 추가하는 방식으로, 가장 많이 사용되고 있는 모델 중 하나이다. 이는 합성 소수 샘플링 기술로 다수 클래스를 샘플링하고 기존 소수 샘플을 보간(Interpolation)하여 새로운 소수 인스턴스를 합성해낸다. 단, 본 연구에서는 이러한 SMOTE를 활용한 Borderline-SMOTE 방법을 적용하였다 [47]. SMOTE 알고리즘이 단순히 minority class에서 랜덤하게 샘플링 했다면, Borderline-SMOTE는 Figure 3.5처럼 다른 class와의 경계에 있는 샘플의 수를 증가시켜서 분류하기 어려운 부분에 집중했다.

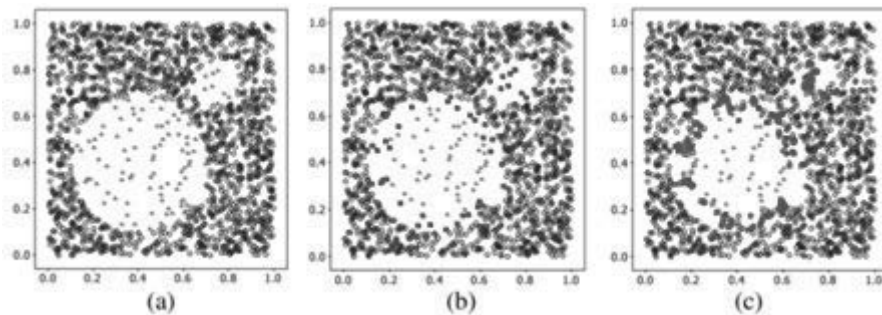


Figure 3.5 Borderline-SMOTE [47].

- (a) The original distribution of Circle data set. (b) The borderline minority examples (solid squares). (c) The borderline synthetic minority examples (hollow squares)

네 번째는 기계학습 모델 선정이다. 우선 전통적인 기계학습 모델 중 Logistic Regression Classifier를 Baseline으로 두고, 추가로 텍스트 분류 시 많이 활용되는 Support Vector Machine(SVM), Naive Bayes(NB) 그리고 Logistic Regression(LR)를 사용하였다. 이러한 전통적인 기계학습 모델의 분류 성능과 인공 신경망 기반의 딥러닝 모델과의 감성 예측 성능을 비교해보는 방법으로 진행하였다. 로지스틱 회귀(Logistic Regression)을 Baseline Model로 선정하게 된 이유는, 본 연구의 감성 분석은 결국 긍정(1)과 부정(0)으로 이진 분류(Binary Classification)하는 문제이기 때문이다. 실제로 문서 분류에 로지스틱 회귀 모델을 많이 적용되는 모델은 아니지만, 이진 분류 문제를 풀기 위한 대표적인 알고리즘으로 로지스틱 회귀를 Baseline Model로 선택하였다. 그리고 총 4가지 모델을 적용하였는데, 첫 번째 모델은 Support Vector Machine(SVM)이다. Support Vector Machine은 지도학습 모델의 일종으로, 클래스 사이의 Margin을 최대화하는 초평면(Hyperplane)을 찾는 방법으로, 해당 초평면에 걸쳐서 이를 지지하는 관측치 들을 Support Vector라고 하여 Support Vector Machine이라고 부른다. SVM은 커널트릭(kernel trick)을 사용하여 주어진 데이터를 고차원 특징 공간(Feature Space)으로 사상(Projection)함으로써, Non-Linear 문제를 Linear 문제로 변환하여 높은 차원의 공간의 데이터를 선형 분리할 수 있다. SVM을 선택한 이유는 다음과 같다. 첫 번째, SVM의 대표적인 장점으로써 고차원의 데이터를 다중공선성 문제(Multicollinearity Problem)를 회피하여 원활하게 분류할 수 있기에, 텍스트와 같은 고차원 인풋에 효과적으로 작동할 수 있다는 판단하였기 때문이다. 본 연구에서는 일반적으로 가장 많이 사용되는 RBF Kernel를 사용하였다. 두 번째 모델은 Naive Bayes(NB)로써, 텍스트 분류를 위해 전통적으로 사용되는 분류기로써 SVM과 더불어 준수한 성능을 보여주는 머신러닝의 주요 알고리즘이다. 지도학습의 일종으로써 베이즈 이론(Bayes theorem)의 확률 모델을 기반으로, 전체 Corpus내 각 단어의 빈도 수를 세어서 위의 사전확률과 우도를 모두 구하는 방법으로 학습을 하고, 사전확률과 우도(Likelihood)의 곱을 계산하는 방식으로 예측한다. 단, 모든 Feature가 서로

독립(independent)이어야 한다는 가정이 필요한데, 즉 단어 간 서로 독립이라는 가정하에 등장확률을 계산한다. 또한 단어의 등장 순서를 무시하고 문서 내 빈도수만을 따져서 문서를 표현하기에 단어 Bag-of-Words 기법과 잘 어울리는 모델이기도 하다. 나이브 베이즈는 적은 계산량으로도 높은 성능을 보이는 장점이 있어, 딥러닝 이전에 텍스트 분류 목적으로 가장 많이 활용되는 대표적인 모델이기에 선택하게 되었다. 세 번째 모델은 Convolutional Neural Network(CNN)으로, 텍스트에 내재된 공간적인 구조 정보를 보존하면서 학습할 수 있는 방법이 필요했으며, 특히 단어나 표현의 등장 순서가 전체적인 텍스트의 의미를 이해하는 데 매우 중요한 정보인 만큼, CNN으로 학습하는 것이 텍스트를 의미적으로 분류하는 데 적절할 것으로 판단했다. 본 논문에서는 각 단어가 벡터로 변환된 문장 행렬을 입력으로 받는 1D CNN 모델을 설계하여 실험을 진행했다. 네 번째는 양방향 LSTM과 어텐션 메커니즘(Bi-LSTM with Attention)이다. 단방향 LSTM으로 텍스트 분류를 수행할 수도 있지만 때로는 양방향 LSTM을 사용하는 것이 더 강력할 수 있으며 여기에 추가적으로 어텐션 메커니즘을 사용할 경우 더욱 효과적인 분류가 가능할 수 있다. 따라서 본 연구에서도 양방향 LSTM과 어텐션 메커니즘으로 문장 단위 감성 분류하기를 수행하였다. 마지막으로 준 지도학습(Semi-Supervised Learning)을 적용하였으며, 준 지도학습의 가장 간단한 방법인 셀프트레이닝(Self-Training)을 활용하였다. 위에서 언급한 기계학습 모델로 Label Data를 학습시켜서 Teacher Model을 구축한다. 그리고 Non-Label 데이터에 대해 Teacher Model로 감성을 예측하고, 이 중 확률이 높은 데이터 위주로 Pseudo-Label을 부여한 뒤 이를 다시 재학습(Retraining)하면서 전체적인 모델의 성능을 향상시키고자 하였다. 이때 확률의 Threshold는 85%로 지정하였는데, 이는 분류기가 85% 이상의 확률로 확신하는 분류 결과만 Label를 부여하겠다는 의미이며, 품질 높은 데이터만 선별하여 모델의 전체 성능을 더욱 높이하고자 하였다. 이렇게 준지도학습을 적용하게 된 가장 큰 이유는, 상대적으로 긍정으로 덜 편향된 Youtube Data(Non-label 데이터)도 함께 학습 데이터로 활용함으로써 자동차

포럼 상 사용자 리뷰 (Label 데이터) 데이터의 데이터 불균형(Data Imbalance) 문제를 조금이나마 해소하고자 하였으며, 보다 더 많은 데이터를 학습시켜서 보다 더 견고한 모델을 기대했기 때문이다.

다섯 번째, 절단된 리뷰(Truncated Text)를 학습한 모델 간 성능을 비교하였고, 가장 나은 모델에 대하여 준지도학습을 통해 분류 성능을 개선하였다.

마지막으로, 지금까지 학습된 모델에 Aspect가 포함된 문장을 인풋으로 넣어 감성을 예측하였다. 즉, 리뷰의 도입부(150개 단어) 부분을 학습한 모델에, 문장 단위 텍스트를 입력하여 감성을 예측하는 것이다. 다만 학습 데이터가 충분하지 않아 문장 단위 감성 분류가 제대로 되지 않을 경우를 대비하여, 감성 분류는 아래 3.3.2.2와 같이 Lexicon Based 감성 분류도 병행하였다.

3.3.2.2 Lexicon 기반 감성 분석

본 연구의 Machine Learning Approach의 결과와 비교해보기 위해 Lexicon based 기반 문장 감성 분류를 병행했으며, TextBlob API에서 제공하는 모듈(Sentiment.Polarity)을 사용하여 각 Aspect-level Sentence의 감성 점수를 산출했다. Textblob은 NLTK를 기반으로 하여 텍스트 처리를 수월하게 할 수 있도록 다양한 기능을 많이 포함하고 있다. "Simplified Text Processing"을 모토로 TextBlob 객체를 생성시키면 주요 메서드를 통해서 텍스트 처리 작업이 굉장히 단순해진다.

3.3.3 A Framework for UX Analysis

이로써 본 연구에서 제안하는 사용자 경험 분석을 위한 ABSA 기반 Framework는 다음과 같다. 먼저 요소 추출(Asspect Extraction)에서는 비지도적인 추출적 접근 방법인 TextRank와 Naïve Method를 적용하여 명사구와 연어를 추출하고, 이를 Human Factor 관점에서 카테고리화 한다. 그리고 감성 분석(Sentiment Analysis)에서는 기계학습 기반의 감성분석과 Lexicon 기반

감성분석을 병렬적으로 진행하여 두 결과를 서로 비교한다. 기계학습 기반의 감성 분석 단계에서는 먼저 Label이 있는 리뷰의 첫 한 두문장만을 잘라낸 Truncated Text를 모델이 학습하게끔 하며, 이때 모델은 텍스트 분류에서 적합한 SVM, NB 및 신경망 모델을 사용한다. 그리고 학습된 모델을 활용하여 Non-Label 데이터에 대해 Label을 부여하는데, 준지도학습의 한 형태인 Self-Training 방법을 사용한다. 즉 Non-Label 데이터에 대해 높은 확률을 갖는 데이터에 대해서 Label을 부여하고, 이를 N차 재학습하여 전체적인 모델의 성능이 높아지도록 한다. 그리고 위에서 추출한 Aspect가 포함된 문장을 불러와 학습된 모델로 긍정/부정 감성을 예측하여 Aspect 단위 감성 분석을 진행한다. Figure 3.6은 본 논문에서 제시하는 사용자 경험 분석 Framework를 도식화한 그림이다

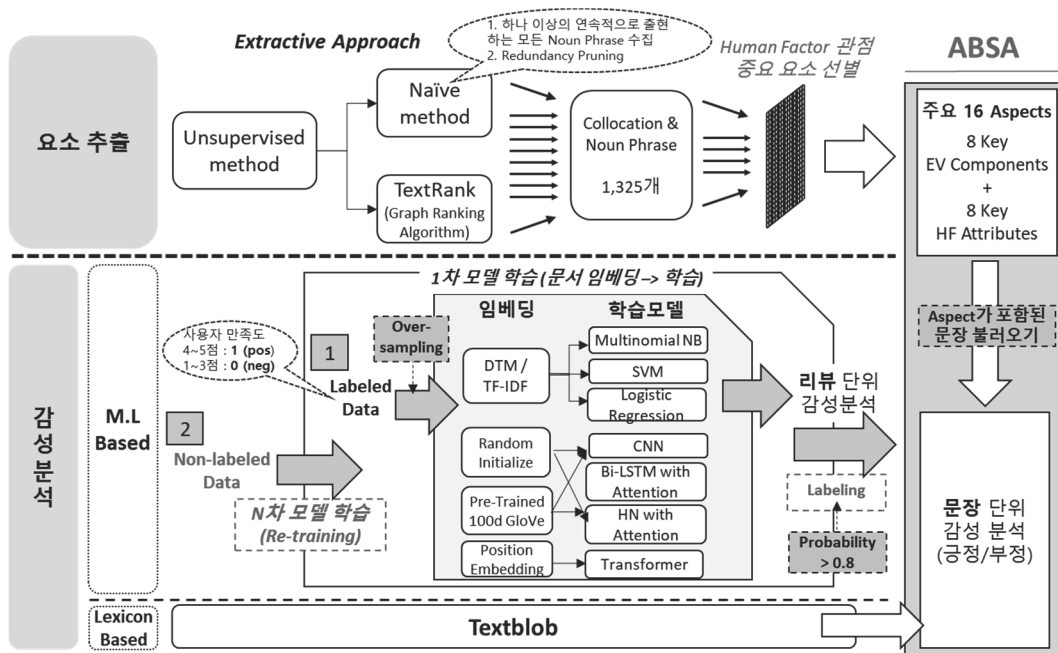


Figure 3.6 A Brand-new Framework: Analysis of User Experience

제 4 장 연구 결과

4.1 요소 추출(Aspect Extraction) 결과

Table 4.1 요소 추출 결과

<Top 8 Components>										EV Characteristics (단위 : 출현 빈도수)									
Charge		Seat		Battery		Room/Space		User Interface		Body/Internals		Design		Utility/SW					
Charge	1052	Seat	359	Battery	988	Room	198	Screen	197	Engine	363	Interior	337	Autopilot	151				
Charger	245	Driver Seat	21	Battery Range	56	Space	188	Display	65	Door	175	Design	177	Software	137				
Super-Charger	66	Back Seat	17	Battery Pack	39	Trunk	125	Pedal	58	Motor	175	Style	132	Cruise Control	128				
Level n Charger	53	Passenger Seat	9	Battery degradation	36	Cargo Space	56	Button	53	Transmission	102	Size	116	Sport Mode	87				
Recharge	16	Seat Heater	6	Battery Capacity	33	Storage	46	Access	46	Suspension	82	Look	134	Navigation	65				
Charge Port	10	Seat Bit	6	Battery Life	21	Leg Room	35	Wind Shield	44	Tire	72	Color	54	Sound Sys.	61				
		Seat Adjustment	5			Head Room	26	Dash board	31	Body	43	Exterior	23	Assist	41				
						Luggage	11	UI	27	Powertrain	27	Appearance	13	Infotainment	35				
						Room front	7	Footprint	22	Hood	24								
								Shifter	9										

<Main 8 Attributes>																			
Driving		Effectiveness		Efficiency		Comfort		Reliability		Cost		Ergonomics		Human Factors					
Acceleration	410	Power	463	Mph	163	Noise	74	Quality	228	Price	460	Ergonomic	52	Safety	182				
Driving	300	Performance	319	Efficiency	88	Comfort	72	Reliability	78	Cost	320	Foot / Leg	57	Luxury	135				
Braking	183	Mileage	290	Capacity	73	Rest	37	Center Gravity	37	Money	303	Head	67	Visibility	88				
Handling	135	Speed	235	Charge Time/Rate	26	Convenience	22	Ride Quality	18	Dealership	221	Seating/Driving Position	11	Attention	55				
Steering	45	Torque	143	Capability	26	Quietness	23	Balance	16	Warranty	205	Knee	6	Responsive-ness	47				
		Fast	73	Lag	12	Family Car	9	Vibration	9	Value	182	Arm	5	Ease of Use	45				
		Mile Range	73							Tax Credit	120								

<Consumer Needs>										<Context of Use>									
Problem	Issue	Need	Trouble	Lack	Com-plaint	Anxiety	Concern	Improv-ement	Draw-back	Limit-ation	Claim	Com-mute	High-way	Ice	Winter	Sum-mer	Snow	Turn	Curve
257	284	90	78	52	45	41	22	27	19	6	8	366	267	187	175	115	79	40	21

요소 추출(Aspect Extraction) 결과 다양한 전기차의 구성 부품 요소와 사용자 특성 요소들이 추출되었음을 확인했다. 표 4.1은 추출된 전기차 주요 구성 요소(Component)들을 항목 별로 카테고리화 하고, 특히 주요 특성(Attribute)들은 효과성(Effectiveness), 효율성(Efficiency), 사용성(Usability) 등의 Human Factor적으로 중요한 항목 중심으로 카테고리화하여 총 8개의 Component와 8개의 Attribute 항목으로 재구성한 결과이다.

본 연구는 대상 제품이 전기차이기 때문에 무엇보다도 전기차만의 특징이 잘 추출될 수 있기를 기대하였는데, Table 4.1에서 볼 수 있듯이 “Charge”, “Battery”,

“Room”, “Room Front”, “Software”, “Motor”등 전기차만이 갖고 있는 제품 하위 요소이거나 또는 차별성을 드러내는 구성 요소들이 추출되었음을 알 수 있다. 또한 “Acceleration”, “Responsiveness”, “Quietness” 등 전기차 만의 갖는 특징인 “반응성(Responsiveness)” 및 “조용함(Quietness/Silence)”과 같은 특성이 Attribute으로 추출되었음을 확인할 수 있다. 또한 “Issue”, “Problem”, “Need”, “Trouble”, “Lack”, “Complain”등 사용자의 불만 사항 및 니즈를 직접적으로 추출할 수 있는 단어들도 상당수 얻을 수 있었다. 마지막으로 전기차를 사용하는 환경(Environment) 또는 사용 맥락(Context of Use) 관련 어휘들이 많이 추출되었는데, 이중 “Ice”, “Winter”, “Snow”, “Cold”와 같은 어휘의 빈도수가 높은 것으로 보아 사용자들이 추운 날씨에 전기차 사용 관련 이슈 사항이 많이 있음을 확인할 수 있다.

4.2 감성 분석(Sentiment Analysis)을 위한 모델링 결과

4.2.1 기계학습 기반 감성 분류 결과에 대한 모델 별 성능 비교

총 6개의 모델을 적용하였으며, 리뷰 단위로 긍정/부정 Label이 있는 데이터를 학습한 감성 분류 모델링 결과는 Table 4.2와 같다. 모델간 성능을 비교하는 기준 지표는 Macro average F1-Score를 선택하였으며 선택 사유는 다음과 같다. 본 연구는 부정적 의견(Minority Class)에 대한 예측하는 성능을 높이는 것이 무엇보다 중요하다. 사용자 리뷰처럼 긍정으로 상당히 편향된 Imbalanced Data를 그대로 학습하는 경우, 모델의 예측 결과가 모두 긍정으로 분류해버릴 가능성이 높다. 이 경우 정확도(Accuracy)는 높게 나오지만 재현율(Recall)이 급격하게 떨어지는 현상이 발생할 수 있다. 단, Macro Average F1-Score의 경우 Precision과 Recall의 산술 평균을 취하여 계산하므로 Minority Class의 재현율(Recall)에 가장 민감하게 반응하는 지표이다. 즉, Macro average F1-Score가 높다는 것은 부정적 의견에 대한 모델의 재현율(Recall)이 높다는 의미이다. 결국 본 연구에서는 Minority Class(Negative Sentiment)를 잘 분류해내는 것이 좋은 모델의 기준인 만큼, Weighted

Average가 아닌 산술평균으로써 Minority Class 분류 결과에 따라 가장 민감하게 반응할 수 있는 Macro F1 Score가 성능 척도로 가장 적합하다고 판단하였다.

결과적으로 CNN이 본 데이터셋에 대해 가장 좋은 성능을 나타내는 모델로 확인되었다. 다른 Attention 기반의 두 모델인 양방향 LSTM(Bi-LSTM with Attention)과 계층적 네트워크(Hierarchical Attention Network)의 경우 예상보다 좋은 결과를 나타내지 못했는데, 이는 데이터셋이 충분하지 못하여 모델이 단어 간의 배열 순서 정보나 단어-문장으로 이어지는 계층 구조에 대한 학습을 충분히 하지 못한 것이 원인으로 추정된다. 아래 Table 4.2는 모델 별 성능 비교 결과이다.

Table 4.2 감성 분석 실험 결과

▪ **성능 지표: Macro average F1-Score**

Model	Embedding	Accuracy	Precision	Recall	F1-score
CNN	GloVe	0.89	0.77	0.80	0.78
Bi-LSTM with attention	Random Initialized	0.85	0.68	0.80	0.72
HN with attention	GloVe	0.85	0.64	0.66	0.65
Transformer	Token & Position Embedding	0.83	0.67	0.73	0.70
Multinomial NB	TF-IDF	0.89	0.72	0.80	0.75
LR		0.92	0.80	0.74	0.76
SVM		0.92	0.80	0.75	0.77

4.2.2 준지도학습(Semi Supervised Learning) 실험 결과

실험 결과, 아래와 같이 준지도학습(Semi Supervised Learning)을 통한 모델의 성능 향상을 확인하였다. Figure 4.1은 표 4.3에서 분류 성능이 가장 좋은 CNN 모델을 기반으로 Non-Label 데이터인 Youtube Comments를 활용한 준지도학습(Semi-Supervised Learning) 실험 결과이다. Figure 4.1의 왼쪽(좌) Matrix는 준지도학습 전의 Label 데이터를 학습한 CNN 모델의 테스트 결과이며, Figure 4.1의 오른쪽(우) Matrix는 준지도학습을 마친 CNN 모델의 테스트 결과이다. 준지도학습은 Label 데이터를 학습한 모델이 Non-Label 데이터를 분류할 때, 분류 확률이 높은 데이터 위주로 Pseudo Label을 부여하며, Pseudo Label가 부여된 데이터를 포함한 전체 데이터셋을 재학습(Retraining)하면서 전체적인 모델의 성능을 개선시키는 Self-Training 방법을 사용했다. 테스트 결과, Loss 값은 상승한 반면, Minority와 Majority 모두 분류 성능이 개선됨으로써 F1 Score와 Accuracy가 모두 0.1씩 상승함을 확인함으로써 준지도학습을 통한 성능 개선 효과를 입증하였다. 향후 더 많은 양질의 Non-Label 데이터를 얻는다면 큰 폭의 성능 개선이 가능할 것으로 보인다.

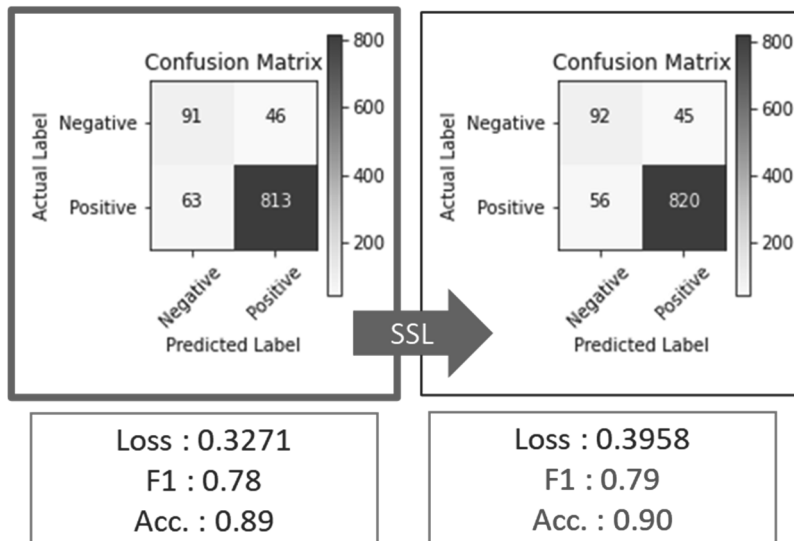


Figure 4.1 준지도학습 실험 결과

4.3 Aspect-Based Sentiment Analysis 결과

4.3.1 기계학습 모델 기반 ABSA 결과

Figure 4.2는 위 Table 4.1에 나타난 추출된 주요 요소(Aspect)들을 포함한 문장에 대하여 chapter 4.2.2의 결과 모델로 문장 단위 감성 분석을 진행한 결과이다. Figure 4.2에서 볼 수 있듯이, 각 Aspect별 출현 빈도를 확인할 수 있으며, 기계학습 모델이 예측한 긍정/부정 클래스의 비율을 확인할 수 있다. 대체적으로 긍정적인 의견이 많으나, 한 쪽으로 완전히 편향되지 않은 결과를 나타냈다. 부정 비율 40%를 기준으로, 전기차 사용자들은 Acceleration(12%), Safety(14%), Visibility(18%), Room(27%), Interior(29%), Reliability(33%), Power(34%)에 대해 긍정적인 의견이 다수인 반면, Noise(40%), Ice(41%), Price(44%), Winter(46%), Seat(52%), Battery(54%)에 대해 다소 부정적인 의견이 많음을 확인할 수 있다. 또한 긍정 의견이 압도적인 온라인 리뷰에 내재된 많은 수의 부정 의견을 발견할 수 있음을 확인할 수 있다.

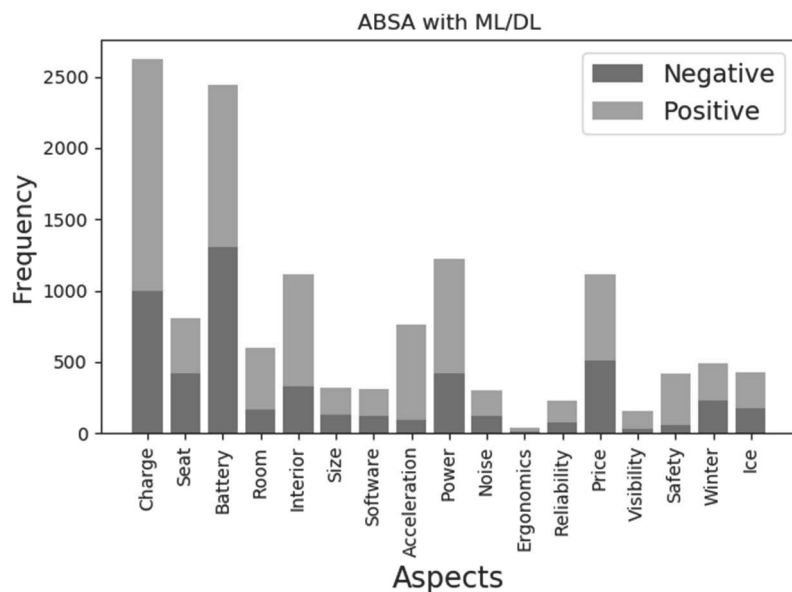


Figure 4.2 기계학습 모델(CNN) 기반 ABSA 결과

한편, 문장 별로 정답 Label이 없는 본 실험의 데이터셋에 대하여, 기계학습 모델이 각 Aspect에 대한 감성 분류(Asspect를 포함한 문장의 긍정/부정 분류) 정확도를 확인하기 위해서는, 직접 휴먼 라벨링(Labeling)을 통해서만 판단할 수 있다. 따라서 본 실험에서는 Sample의 수가 상대적으로 적은 “Ergonomic”를 대상으로 분류 성능을 확인해보았다. “Ergonomic”가 포함된 문장 총 51개에 대해, 모델이 예측한 결과는 긍정 의견 38개, 부정 의견 13개이며(상세: Appendix I-1), 그리고 Table 4.3은 휴먼 라벨링을 통해 모델의 분류 정확도를 평가한 결과이다. “Ergonomics”가 포함된 문장 Level 감성 분류 정확도는 약 92.1%으로 상당히 높았다 (상세 Appendix I-2). 이는 리뷰의 도입부 한 두 문장을 학습한 기계학습 모델로 Aspect-level의 문장 단위 감성 분류를 진행한 방법이 타당함을 보여준다.

Table 4.3 딥러닝 모델의 문장 분류 성능 실험

Confusion Matrix		Predict	
		Pos.	Neg.
Actual	Pos.	35	1
	Neg.	3	12

4.3.2 Lexicon 기반 ABSA 결과

Figure 4.3는 위 Table 4.1에 나타난 주요 Aspect들이 포함된 문장을 Lexicon 기반으로 문장 단위 감성 분류한 결과이다. 이 경우 문장 별 Negative(-1)와 Positive(1) 사이의 감성 극성 점수 계산이 가능하다. Figure 4.3. Boxplot(上)을 확인해보면, Aspect별 사용자의 긍정/부정 감성의 밀집도 및 편차를 확인할 수 있다. 아래 Figure 4.3. Scatterplot(下)을 통해 Aspect별 감성 분포를 확인할 수 있는데, 여기서 한 개의 점이 Aspect가 포함된 문장이며, 0을 기준으로 그 이하에 분포가 높을 수록 사용자의 부정적인 의견이 상대적으로 많은 것을 의미한다. 사용자들의 Room, Interior, Acceleration, Visibility, Safety에 대해서는 긍정적인 의견이 많은 반면,

Charge, Battery, Power, Ice, Noise에 대해서는 부정적 의견이 많음을 확인할 수 있다.

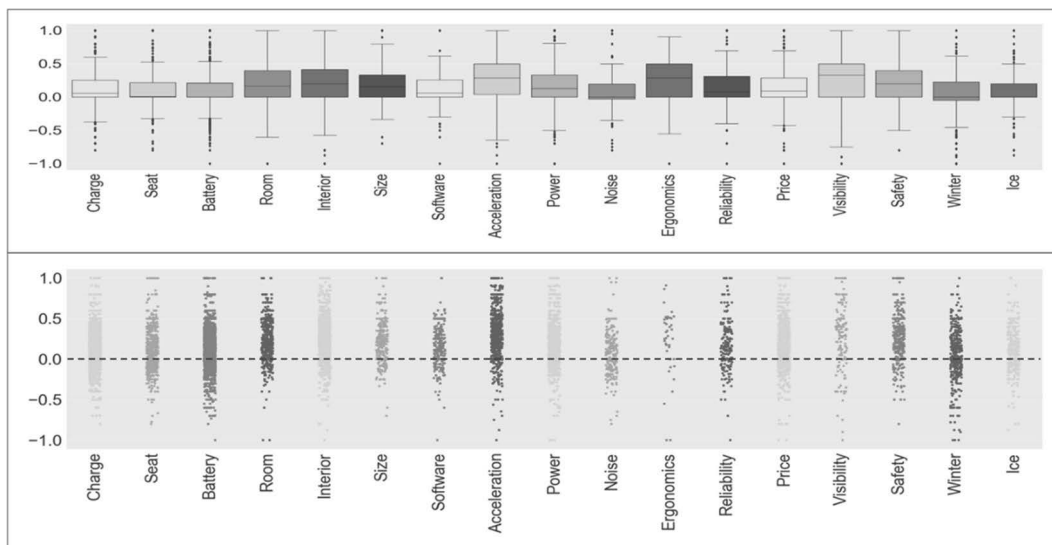


Figure 4.3 Lexicon 기반 ABSA 결과 I: Boxplot(上) / Scatterplot(下)

Figure 4.4는 Lexicon 기반의 분류 결과를 Figure 4.2와 비교하기 위해 동일한 양식으로 시각화한 것이며, 두 분포를 비교한 결과는 Table 4.4와 같다.

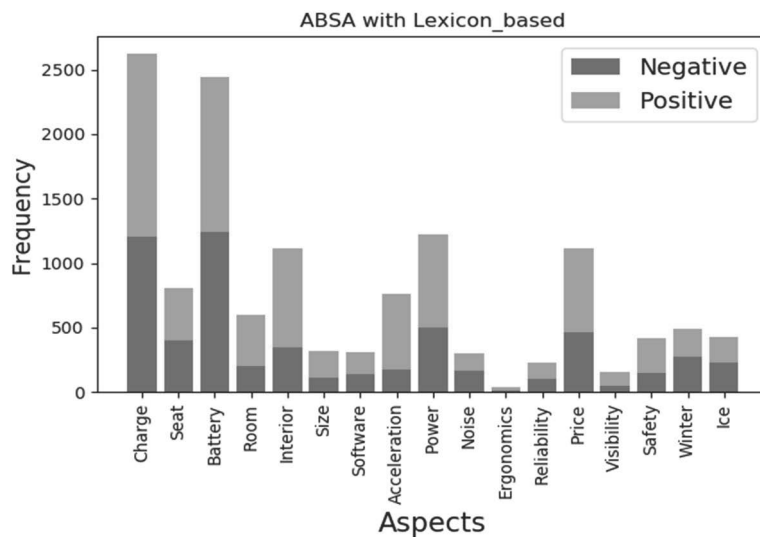


Figure 4.4 Lexicon 기반 ABSA 결과 II

결과적으로 총 17개 항목 중 3개 항목 (Noise, Winter, Ice)을 제외한 14개 항목에서 긍정/부정 예측 방향성이 일치하였다. 여기서 예측 방향성이 일치한다는 의미는 하나의 Aspect에 대한 긍정/부정 감성 분류 비율이 Lexicon 기반 모델과 ML 기반 모델이 서로 유사한 것을 의미한다. (예를 들어 “Noise”의 경우 Lexicon은 부정으로 예측한 반면, ML 기반 모델은 긍정으로 다수 예측하여 예측 방향성이 불일치하였다.) 전반적으로 Lexicon 기반 모델과 ML 기반 모델이 서로 유사한 분류를 하였기 때문에, 본 연구의 기계학습 기반의 모델이 비교적 안정적이라고 할 수 있는 Lexicon 기반 감성 예측 방법과 비슷하게 안정적인 분류 성능을 갖고 있는 점을 간접적으로 확인할 수 있다.

Table 4.4 Lexicon Based 모델과 ML 모델간 문장 긍정/부정 분류 결과 비교

모델	Aspect	Charge	Seat	Battery	Room	Interior	Size	Software	Acceleration	Power	Noise	Ergonomics	Reliability	Price	Visibility	Safety	Winter	Ice
Lexicon	Pos	54	50	49	67	69	65	55	77	61	45	73	55	58	73	66	44	47
	Neg	46	50	51	33	31	35	45	23	39	55	27	45	42	27	34	56	53
ML	Pos	62	48	46	73	71	61	63	88	66	60	78	67	56	82	86	54	59
	Neg	38	52	54	27	29	39	37	12	34	40	22	33	44	18	14	46	41

4.4 사용자 긍/부정 경험에 영향을 미치는 Contributing Factor

특정 Aspect에 대한 차량의 상세 Spec(X)과 사용자의 감성(y) 간의 관계를 상관관계분석과 다중회귀분석을 통해 확인해보았다. Figure 4.5는 “Leg Room”이라는 Aspect에 대한 상세 스펙인 “Front leg room“의 실제 치수(X)와 “Leg Room”에 대한 Lexicon 기반 감성 점수(y) 간의 관계를 나타낸 그래프이다. “Front leg room“의 실제 치수는 Edmunds.com에서 크롤링을 통해 얻었다. Figure 4.5에서 나타나듯이, 명확한 양 또는 음의 상관성은 나타나지 않았으며, 다른 Aspect에도 비슷한 결과를 얻었다. “leg room“과 같이 사용자의 신체와 직접적으로 관련된 요소의 경우, 단편적인 상관관계 분석보다는 운전자의 키, 다리 길이와 같은 신체 정보를 함께 반영했을 시 보다 더 의미 있는 해석이 가능할 수 있을 것으로 예상된다.

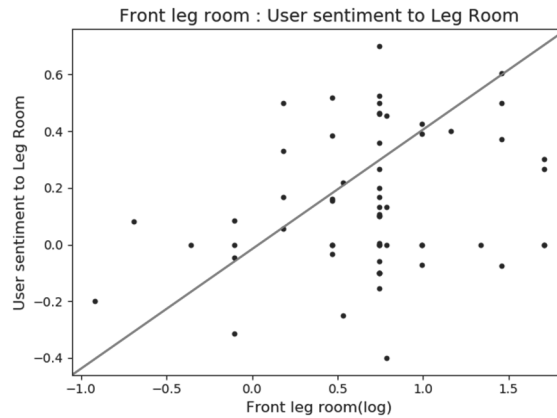


Figure 4.5 Aspect의 차량 별 실제 치수와 사용자 감성 간 상관 관계 예시

Table 4.5는 전기차의 “공간(Room)”에 대해 사용자의 긍정/부정 감성에 영향력 있는 상세 스펙을 확인하기 위해 다중선형회귀분석을 실시한 결과이다. “Room”과 관련된 총 8개, “Front Head Room”, “Front Shoulder Room”, “Front Leg Room”, “Rear Head Room”, “Rear Leg Room”, “Rear Hip Room”, “Front Hip Room”에 대한 상세 스펙을 얻었으며, 이는 Edmunds.com에서 크롤링을 통해 획득하였다.

Table 4.5 “Room”에 대한 사용자 감성 극성과 상세 스펙간 회귀분석 결과

X	F.Head Room	F.Shoulder Room	F. Leg Room	R.Head Room	R. Leg Room	R.Shoulder Room	R. Hip Room	F. Hip Room
coefficient	-0.013	0.021	-0.035	-0.004	-0.008	0.004	0.024	-0.045

R^2 는 0.096으로 의미 있는 결과는 얻지 못하였으며, Room 이외의 다른 Aspect들에 대해서도 비슷한 결과를 얻었다.

결과적으로 주요 Aspect에 대한 사용자 감성에 영향력 있는 변수 및 명확한 (양/음) 상관성을 띄는 변수를 찾기 어려웠는데, 이는 현재 차종/데이터의 수가 많지 않은 점이 원인이 될 수 있다. 또한 다른 외부 변수, 예를 들어 사용자의 신체 정보를 얻을 경우 관계성을 찾아 더욱 의미 있는 해석이 가능할 수 있다.

제 5 장 결론

5.1 결론(Conclusion)

본 논문에서는 사용자 경험을 효과적으로 분석하기 위한 Aspect-Based Sentiment Analysis 기반의 새로운 사용자 경험(UX) 분석 Framework를 제안하였다. 이를 통해 전기차를 대상으로 사용자 경험을 분석하면서 얻은 몇 가지 발견은 다음과 같다. 첫 번째, 전기차의 사용자 리뷰로부터, 형용사와 명사 중심의 명사구 추출을 통해 차량 요소를 추출한 결과, 전기차의 다양한 구성 요소(Component)들 뿐만 아니라, Human Factor 특성(Attribute)를 추출할 수 있음을 확인하였다.

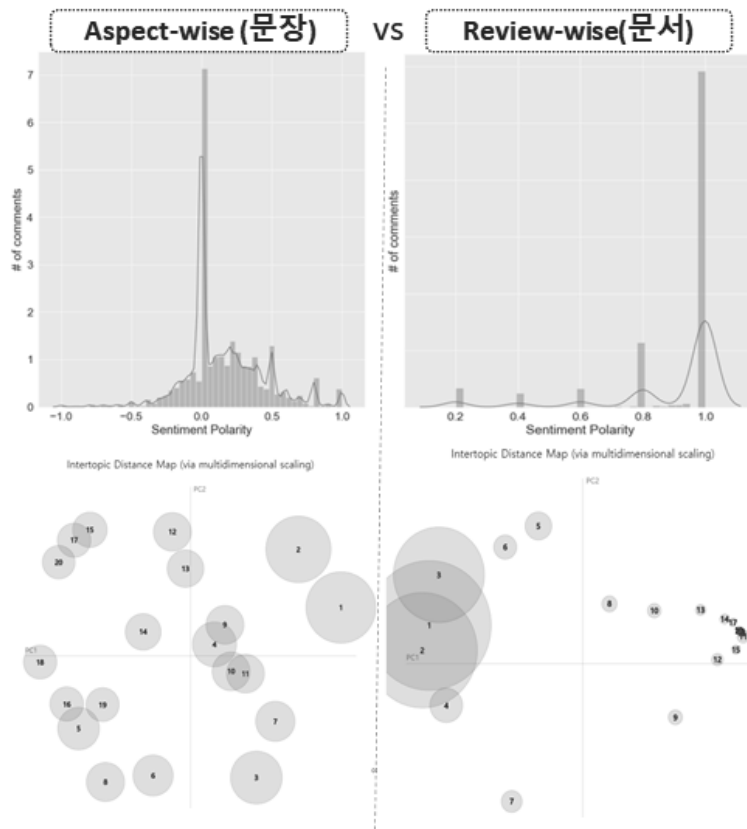


Figure 5.1 Aspect-wise 분석과 Review-wise 분석 비교

두 번째, Aspect 단위의 감성분석을 통해 Review 단위 감성분석 대비 더욱 다양한 의견을 얻을 수 있음을 확인하였다. Figure 5.1에서 확인할 수 있듯이, Review-wise 분석 시, 토픽 클러스터링의 분포가 굉장히 한쪽으로 뭉쳐져 있는데 이는 긍정에 높은 의견을 반영하는 것으로 유추해볼 수 있다. 반면 Aspect-wise 분석 시, 다양한 토픽 클러스터링의 분포를 확인할 수 있었다.

세 번째, Minority Class에 대해 오버샘플링(Over-sampling)을 통해 데이터 불균형(Data Imbalance) 문제를 극복할 수 있음을 확인하였다. (Appendix II)

네 번째, 준지도학습(Semi Supervised Learning)을 통한 분류 성능 개선을 확인함으로써 Label이 없는 데이터에 대한 활용 방법을 제시하였다.

다섯 번째, 리뷰의 도입부 감성 극성이 전체 리뷰의 감성을 결정한다는 전제 하에, 절단된 리뷰 학습을 통한 문장 분류가 효과적일 수 있음을 확인하였다. 특히 딥러닝 모델을 통해 단어의 의미에 따른 긍/부정 분류 뿐만 아니라, 문맥적인 의미 파악을 통한 감성 분류 또한 가능함을 확인하였다 (상세 Appendix I-2).

마지막으로 기계학습을 이용한 Aspect-Based Sentiment Analysis 기반으로 UX를 분석하여, 전기차의 긍정적 요소와 부정적 요소를 확인할 수 있었다. 특히 부정적 요소의 경우 동시 출현(Co-occurrence) 빈도수 기반 연관 어휘와 함께 해석을 통해 사용자의 니즈를 발굴할 수 있었다. 사용자는 Charge 효율성 문제, 좁은 뒷자리(Back Seat), 추운 겨울 날씨에 배터리 기능 저하 문제, 그리고 조용한 환경에 따른 외부 소음 문제에 대한 개선 니즈가 있음을 확인할 수 있었다.

<div>+</div> 전기차의 긍정적 요소	<div>-</div> 전기차의 부정적 요소
<ul style="list-style-type: none"> Acceleration / Room / Interior / Power / Visibility / Safety / Ergonomics 	<ul style="list-style-type: none"> Charge, Seat, Battery, Noise, Winter <ul style="list-style-type: none"> Charge Rate / Time & Miles per charge 문제 Back Seat의 협소함 추운 날씨의 Battery 기능 저하 및 고갈에 따른 Mile range 감소 문제 내부 (Suspension 등) Noise, Wind/Road Noise가 오히려 부각되는 사항

Figure 5.2 전기차 긍정/부정 요소 및 도출된 사용자 니즈

5.2 연구 기여(Contribution)

제품의 주요 Aspect를 추출하고, 추출된 Aspect 별로 사용자 긍정/부정 감성 분포를 통해 부정적인 의견(사용자 니즈)이 많은 Aspect를 한 눈에 쉽게 확인할 수 있는 새로운 UX 분석 프레임워크를 개발하였다. 또한 Aspect에 대해 ML 또는 Lexicon 기반 (긍정/부정) 감성 예측 결과를 지지하는 문장들을 모두 불러와, 문장 별 감성 분석 및 사용 목적에 맞게 추가적인 분석을 진행할 수 있는 시스템을 개발하여, 세분화된 사용자/고객 조사가 필요한 관련 분야에 기여하였다.

5.3 한계점(Limitation)

Supervised Learning 기반의 문장의 감성 예측 모델링을 구현하고자 하였으며, 다만 문장 단위의 감성 Label이 없기 때문에, 사용자 리뷰 서두의 1~2문장을 학습한 모델을 활용하여 문장단위 감성분석을 진행하였다. 본 방법이 데이터는 리뷰의 개수가 상당히 많을 경우 예측 정확도가 더욱 높을 것으로 예상된다. 따라서 본 연구의 전제 조건인 Label 데이터가 부족한 상황에서는 양질의 Non-label 데이터의 수를 많이 얻을 수 있다면 더욱 좋은 성능의 모델을 구성할 수 있을 것으로 보인다. 이러한 측면에서 Non-label 데이터 중 UX를 나타내는 텍스트만 선별해내는 부분이 더욱 강화된다면, 더 나은 모델을 만들 수 있을 것이다.

또한 심화 연구(영향력 있는 차량 스펙 조사)의 경우, 사용자의 감성에 영향력 있는 차량 세부 스펙 또는 양/음의 상관성을 갖는 차량 스펙이 무엇인지 찾기 어려웠는데, 이는 얻을 수 있는 데이터의 변수가 제한적이기 때문이다. 외부 변수로 사용자의 키와 몸무게, 팔다리 길이 등의 신체적 조건을 독립 변수로 활용 가능할 때 더욱 의미 있는 결과를 얻을 수 있을 것으로 예상된다.

Appendix

Appendix I-1. “Ergonomics”에 대한 Aspect 별 Opinion Mining

```

1. "positive": [
2.   "I like it.",
3.   "It's great.",
4.   "I love it.",
5. ],
6. "neutral": [
7.   "It's OK.",
8.   "It's fine.",
9. ],
10. "negative": [
11.   "I don't like it.",
12.   "I hate it.",
13.   "It's terrible.",
14.   "I don't want it.",
15. ],
16. "mixed": [
17.   "I like some parts, but not others.",
18.   "It's a mixed bag.",
19. ],
20. "uncertain": [
21.   "I'm not sure.",
22.   "I'm not certain.",
23. ],
24. "other": [
25.   "I don't know.",
26.   "I'm not sure.",
27. ],
28. ]
29.
30. # Example 2: A list of sentences expressing various emotions and opinions.
31. # This list is more diverse and includes more complex sentences.
32. # It also includes some sentences that are more specific and detailed.
33. # The sentences are grouped into the same categories as the first list.
34. # The categories are: positive, neutral, negative, mixed, uncertain, and other.
35. # The sentences are as follows:
36.
37. # Positive sentences:
38. "I love this car! It's so fast and powerful!"
39. "The engine is amazing! It's so smooth and quiet."
40. "The interior is so comfortable and spacious."
41. "The exterior is so sleek and modern."
42. "The overall design is so impressive."
43. "I'm really impressed with the quality of the materials."
44. "The car is so easy to drive and handle."
45. "The ride is so smooth and comfortable."
46. "The car is so reliable and durable."
47. "The overall performance is so excellent."
48. "I'm really happy with my purchase."
49. "The car is so much better than the others."
50. "The car is so much more fun to drive."
51. "The car is so much more comfortable."
52. "The car is so much more reliable."
53. "The car is so much more durable."
54. "The car is so much more powerful."
55. "The car is so much faster."
56. "The car is so much more agile."
57. "The car is so much more maneuverable."
58. "The car is so much more stable."
59. "The car is so much more secure."
60. "The car is so much more comfortable."
61. "The car is so much more reliable."
62. "The car is so much more durable."
63. "The car is so much more powerful."
64. "The car is so much faster."
65. "The car is so much more agile."
66. "The car is so much more maneuverable."
67. "The car is so much more stable."
68. "The car is so much more secure."
69. "The car is so much more comfortable."
70. "The car is so much more reliable."
71. "The car is so much more durable."
72. "The car is so much more powerful."
73. "The car is so much faster."
74. "The car is so much more agile."
75. "The car is so much more maneuverable."
76. "The car is so much more stable."
77. "The car is so much more secure."
78. "The car is so much more comfortable."
79. "The car is so much more reliable."
80. "The car is so much more durable."
81. "The car is so much more powerful."
82. "The car is so much faster."
83. "The car is so much more agile."
84. "The car is so much more maneuverable."
85. "The car is so much more stable."
86. "The car is so much more secure."
87. "The car is so much more comfortable."
88. "The car is so much more reliable."
89. "The car is so much more durable."
90. "The car is so much more powerful."
91. "The car is so much faster."
92. "The car is so much more agile."
93. "The car is so much more maneuverable."
94. "The car is so much more stable."
95. "The car is so much more secure."
96. "The car is so much more comfortable."
97. "The car is so much more reliable."
98. "The car is so much more durable."
99. "The car is so much more powerful."
100. "The car is so much faster."
101. "The car is so much more agile."
102. "The car is so much more maneuverable."
103. "The car is so much more stable."
104. "The car is so much more secure."
105. "The car is so much more comfortable."
106. "The car is so much more reliable."
107. "The car is so much more durable."
108. "The car is so much more powerful."
109. "The car is so much faster."
110. "The car is so much more agile."
111. "The car is so much more maneuverable."
112. "The car is so much more stable."
113. "The car is so much more secure."
114. "The car is so much more comfortable."
115. "The car is so much more reliable."
116. "The car is so much more durable."
117. "The car is so much more powerful."
118. "The car is so much faster."
119. "The car is so much more agile."
120. "The car is so much more maneuverable."
121. "The car is so much more stable."
122. "The car is so much more secure."
123. "The car is so much more comfortable."
124. "The car is so much more reliable."
125. "The car is so much more durable."
126. "The car is so much more powerful."
127. "The car is so much faster."
128. "The car is so much more agile."
129. "The car is so much more maneuverable."
130. "The car is so much more stable."
131. "The car is so much more secure."
132. "The car is so much more comfortable."
133. "The car is so much more reliable."
134. "The car is so much more durable."
135. "The car is so much more powerful."
136. "The car is so much faster."
137. "The car is so much more agile."
138. "The car is so much more maneuverable."
139. "The car is so much more stable."
140. "The car is so much more secure."
141. "The car is so much more comfortable."
142. "The car is so much more reliable."
143. "The car is so much more durable."
144. "The car is so much more powerful."
145. "The car is so much faster."
146. "The car is so much more agile."
147. "The car is so much more maneuverable."
148. "The car is so much more stable."
149. "The car is so much more secure."
150. "The car is so much more comfortable."
151. "The car is so much more reliable."
152. "The car is so much more durable."
153. "The car is so much more powerful."
154. "The car is so much faster."
155. "The car is so much more agile."
156. "The car is so much more maneuverable."
157. "The car is so much more stable."
158. "The car is so much more secure."
159. "The car is so much more comfortable."
160. "The car is so much more reliable."
161. "The car is so much more durable."
162. "The car is so much more powerful."
163. "The car is so much faster."
164. "The car is so much more agile."
165. "The car is so much more maneuverable."
166. "The car is so much more stable."
167. "The car is so much more secure."
168. "The car is so much more comfortable."
169. "The car is so much more reliable."
170. "The car is so much more durable."
171. "The car is so much more powerful."
172. "The car is so much faster."
173. "The car is so much more agile."
174. "The car is so much more maneuverable."
175. "The car is so much more stable."
176. "The car is so much more secure."
177. "The car is so much more comfortable."
178. "The car is so much more reliable."
179. "The car is so much more durable."
180. "The car is so much more powerful."
181. "The car is so much faster."
182. "The car is so much more agile."
183. "The car is so much more maneuverable."
184. "The car is so much more stable."
185. "The car is so much more secure."
186. "The car is so much more comfortable."
187. "The car is so much more reliable."
188. "The car is so much more durable."
189. "The car is so much more powerful."
190. "The car is so much faster."
191. "The car is so much more agile."
192. "The car is so much more maneuverable."
193. "The car is so much more stable."
194. "The car is so much more secure."
195. "The car is so much more comfortable."
196. "The car is so much more reliable."
197. "The car is so much more durable."
198. "The car is so much more powerful."
199. "The car is so much faster."
200. "The car is so much more agile."
201. "The car is so much more maneuverable."
202. "The car is so much more stable."
203. "The car is so much more secure."
204. "The car is so much more comfortable."
205. "The car is so much more reliable."
206. "The car is so much more durable."
207. "The car is so much more powerful."
208. "The car is so much faster."
209. "The car is so much more agile."
210. "The car is so much more maneuverable."
211. "The car is so much more stable."
212. "The car is so much more secure."
213. "The car is so much more comfortable."
214. "The car is so much more reliable."
215. "The car is so much more durable."
216. "The car is so much more powerful."
217. "The car is so much faster."
218. "The car is so much more agile."
219. "The car is so much more maneuverable."
220. "The car is so much more stable."
221. "The car is so much more secure."
222. "The car is so much more comfortable."
223. "The car is so much more reliable."
224. "The car is so much more durable."
225. "The car is so much more powerful."
226. "The car is so much faster."
227. "The car is so much more agile."
228. "The car is so much more maneuverable."
229. "The car is so much more stable."
230. "The car is so much more secure."
231. "The car is so much more comfortable."
232. "The car is so much more reliable."
233. "The car is so much more durable."
234. "The car is so much more powerful."
235. "The car is so much faster."
236. "The car is so much more agile."
237. "The car is so much more maneuverable."
238. "The car is so much more stable."
239. "The car is so much more secure."
240. "The car is so much more comfortable."
241. "The car is so much more reliable."
242. "The car is so much more durable."
243. "The car is so much more powerful."
244. "The car is so much faster."
245. "The car is so much more agile."
246. "The car is so much more maneuverable."
247. "The car is so much more stable."
248. "The car is so much more secure."
249. "The car is so much more comfortable."
250. "The car is so much more reliable."
251. "The car is so much more durable."
252. "The car is so much more powerful."
253. "The car is so much faster."
254. "The car is so much more agile."
255. "The car is so much more maneuverable."
256. "The car is so much more stable."
257. "The car is so much more secure."
258. "The car is so much more comfortable."
259. "The car is so much more reliable."
260. "The car is so much more durable."
261. "The car is so much more powerful."
262. "The car is so much faster."
263. "The car is so much more agile."
264. "The car is so much more maneuverable."
265. "The car is so much more stable."
266. "The car is so much more secure."
267. "The car is so much more comfortable."
268. "The car is so much more reliable."
269. "The car is so much more durable."
270. "The car is so much more powerful."
271. "The car is so much faster."
272. "The car is so much more agile."
273. "The car is so much more maneuverable."
274. "The car is so much more stable."
275. "The car is so much more secure."
276. "The car is so much more comfortable."
277. "The car is so much more reliable."
278. "The car is so much more durable."
279. "The car is so much more powerful."
280. "The car is so much faster."
281. "The car is so much more agile."
282. "The car is so much more maneuverable."
283. "The car is so much more stable."
284. "The car is so much more secure."
285. "The car is so much more comfortable."
286. "The car is so much more reliable."
287. "The car is so much more durable."
288. "The car is so much more powerful."
289. "The car is so much faster."
290. "The car is so much more agile."
291. "The car is so much more maneuverable."
292. "The car is so much more stable."
293. "The car is so much more secure."
294. "The car is so much more comfortable."
295. "The car is so much more reliable."
296. "The car is so much more durable."
297. "The car is so much more powerful."
298. "The car is so much faster."
299. "The car is so much more agile."
300. "The car is so much more maneuverable."
301. "The car is so much more stable."
302. "The car is so much more secure."
303. "The car is so much more comfortable."
304. "The car is so much more reliable."
305. "The car is so much more durable."
306. "The car is so much more powerful."
307. "The car is so much faster."
30
```

Appendix I-2. “Ergonomics”에 대한 예측 정확도 샘플 검증

- 딥러닝 모델이 분류한 문장의 긍정/부정 분류가 정확히 되었는지 직접 문장을 읽고 Human Labeling을 통해 모델의 분류 정확도 평가

- Aspect : **"Ergonomic"**

Confusion Matrix		Predict	
		Pos.	Neg.
Actual	Pos.	35	1
	Neg.	3	12

- 분류 정확도 측정 결과 **92.1%의 상당히 정확한 분류 성능**을 보임

- 단어 의미 이상으로, 문장 문맥을 파악한 긍/부정 분류 가능

```
'negative': [
    "The interior ergonomics are awful",
    "The ergonomics of this car with everything in this center screen is absolutely horrible and a nightmare",
    "Fantastic economy\n\n\n\nBuying an american car\n\nCons Price even after dealer discount\n\n\n\n\n\n craftsmanship lots of problems\n\n\n\n\n\n Poor ergonomics \n\n",
    "Minor issues but other manufacturers have figured out ergonomics",
    "Fourth Porsche please go pick up a Jeep Grand Cherokee and make it a required course for the folks who design your ergonomics",
    "Cheap iPad with bad ergonomics",
    "Poor ergonomics",
    "Look at the Model is awful interface and ergonomics",
    "Ergonomic nightmare for driver especially if driving long hours I have power driver seat too",
    "I have had Jeep Toyota Nissan MB and other porsche and this is the worst ergonomic design I have ever seen",
    "The seat back has a weird shape it feels as if my shoulders are being pushed forward so I have an ergonomic cushioned frame I bought to make it work",
    "Yes there are a few little ergonomic issues about the car that frustrate me and the steering requires real attention on the freeway but in general for what it is in  

    "If it solved my ergonomic problem I would lease or buy the when my lease expires"
```

Appendix II. SMOTE의 효과성 입증

	precision	recall	f1-score	support
positive	0.00	0.00	0.00	112
negative	0.89	1.00	0.94	901
accuracy			0.89	1013
macro avg	0.44	0.50	0.47	1013
weighted avg	0.79	0.89	0.84	1013

Precision: 0.44471865745310957
 Recall: 0.5
 f1-score: 0.47074190177630453
 Confusion Matrix
 [[0 112]
 [0 901]]
 Accuracy: 0.861796643632774
 Loss: 4.773434812134443

	precision	recall	f1-score	support
positive	0.66	0.44	0.53	112
negative	0.93	0.97	0.95	901
accuracy			0.91	1013
macro avg	0.80	0.70	0.74	1013
weighted avg	0.90	0.91	0.91	1013

Precision: 0.7975347552024068
 Recall: 0.704876526882131
 f1-score: 0.7395278167367929
 Confusion Matrix
 [[49 63]
 [25 876]]
 Accuracy: 0.9200394866732478
 Loss: 2.7617974904195766

	precision	recall	f1-score	support
positive	0.41	0.41	0.41	112
negative	0.93	0.93	0.93	901
accuracy			0.87	1013
macro avg	0.67	0.67	0.67	1013
weighted avg	0.87	0.87	0.87	1013

Precision: 0.6668731563421829
 Recall: 0.6681762327572538
 f1-score: 0.6675205132950829
 Confusion Matrix
 [[46 66]
 [67 834]]



SMOTE Oversampling

	precision	recall	f1-score	support
positive	0.49	0.70	0.57	112
negative	0.96	0.91	0.93	901
accuracy			0.89	1013
macro avg	0.72	0.80	0.75	1013
weighted avg	0.91	0.89	0.89	1013

Precision: 0.7238203399765534
 Recall: 0.8027092912636753
 f1-score: 0.7536974310818969
 Confusion Matrix
 [[78 34]
 [82 819]]

Multinomial Naïve Bayes

	precision	recall	f1-score	support
positive	0.66	0.51	0.58	112
negative	0.94	0.97	0.95	901
accuracy			0.92	1013
macro avg	0.80	0.74	0.76	1013
weighted avg	0.91	0.92	0.91	1013

Precision: 0.8017297609191942
 Recall: 0.7383710559695575
 f1-score: 0.7649028579006697
 Confusion Matrix
 [[57 55]
 [29 872]]

Logistic Regression

	precision	recall	f1-score	support
positive	0.66	0.54	0.60	112
negative	0.94	0.96	0.95	901
accuracy			0.92	1013
macro avg	0.80	0.75	0.77	1013
weighted avg	0.91	0.92	0.91	1013

Precision: 0.800239597942964
 Recall: 0.7545633819565563
 f1-score: 0.7747712996075595
 Confusion Matrix
 [[61 51]
 [32 869]]

SVM

참고 문헌

- [1] Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, 52(10), 144-147.
- [2] 친환경차보급로드맵, <http://www.me.go.kr/home/web/board/read.do?boardMasterId=1&boardId=935880&menuId=286>
- [3] Executive Summary Electric Vehicle Outlook, Bloomberg New Energy Finance's (2017), https://data.bloomberglp.com/bnef/sites/14/2017/07/BNEF_EVO_2017_ExecutiveSummary.pdf
- [4] Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177).
- [5] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., ... & Hoste, V. (2016, June). Semeval-2016 task 5: Aspect based sentiment analysis. In *10th International Workshop on Semantic Evaluation (SemEval 2016)*.
- [6] Thet, T. T., Na, J. C., & Khoo, C. S. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of information science*, 36(6), 823-848.
- [7] Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., & Androutsopoulos, I. (2015, June). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)* (pp. 486-495).
- [8] Manandhar, S. (2014). Semeval-2014 task 4: Aspect based sentiment analysis. In

Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014).

- [9] Etzioni, O., Cafarella, M., Downey, D., Popescu, A. M., Shaked, T., Soderland, S., ... & Yates, A. (2005). Unsupervised named-entity extraction from the web: An experimental study. *Artificial intelligence*, 165(1), 91-134.
- [10] Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Systems with Applications*, 118, 272-299.
- [11] Wang, H., Lu, Y., & Zhai, C. (2010, July). Latent aspect rating analysis on review text data: a rating regression approach. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 783-792).
- [12] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- [13] Wang, B., & Liu, M. (2015). Deep learning for aspect-based sentiment analysis. Stanford University report.
- [14] Poria, S., Cambria, E., Gelbukh, A., Bisio, F., & Hussain, A. (2015). Sentiment data flow analysis by means of dynamic linguistic patterns. *IEEE Computational Intelligence Magazine*, 10(4), 26-36.
- [15] Piryani, R., Gupta, V., & Singh, V. K. (2017). Movie Prism: A novel system for aspect level sentiment profiling of movies. *Journal of Intelligent & Fuzzy Systems*, 32(5), 3297-3311.
- [16] Poria, S., Chaturvedi, I., Cambria, E., & Bisio, F. (2016, July). Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In *2016 international joint conference on neural networks (IJCNN)* (pp. 4465-4473). IEEE.

- [17] Alam, M. H., Ryu, W. J., & Lee, S. (2016). Joint multi-grain topic sentiment: modeling semantic aspects for online reviews. *Information Sciences*, 339, 206-223.
- [18] García-Pablos, A., Cuadros, M., & Rigau, G. (2018). W2VLDA: almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications*, 91, 127-137.
- [19] Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108, 42-49.
- [20] Nguyen, T. H., & Shirai, K. (2015, September). Phrasernn: Phrase recursive neural network for aspect-based sentiment analysis. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 2509-2514).
- [21] Poria, S., Chaturvedi, I., Cambria, E., & Hussain, A. (2016, December). Convolutional MKL based multimodal emotion recognition and sentiment analysis. In *2016 IEEE 16th international conference on data mining (ICDM)* (pp. 439-448). IEEE.
- [22] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104-3112).
- [23] Mihalcea, R., & Tarau, P. (2004, July). Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 404-411).
- [24] Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2017). Sentiment analysis is a big suitcase. *IEEE Intelligent Systems*, 32(6), 74-80.
- [25] Cambria, E; Schuller, B; Xia, Y; Havasi, C (2013). "New avenues in opinion mining and sentiment analysis". *IEEE Intelligent Systems*. 28 (2): 15–21.
- [26] Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016, November). Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical*

methods in natural language processing (pp. 606-615).

- [27] Lei, T., Barzilay, R., & Jaakkola, T. (2016, November). Rationalizing Neural Predictions. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 107-117).
- [28] Ma, D., Li, S., Zhang, X., & Wang, H. (2017, August). Interactive attention networks for aspect-level sentiment classification. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (pp. 4068-4074).
- [29] Tang, D., Qin, B., & Liu, T. (2016, November). Aspect Level Sentiment Classification with Deep Memory Network. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 214-224).
- [30] Liu, J., & Zhang, Y. (2017, April). Attention modeling for targeted sentiment. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers (pp. 572-577).
- [31] Chen, P., Sun, Z., Bing, L., & Yang, W. (2017, September). Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the 2017 conference on empirical methods in natural language processing (pp. 452-461).
- [32] Fan, F., Feng, Y., & Zhao, D. (2018). Multi-grained attention network for aspect-level sentiment classification. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (pp. 3433-3442).
- [33] Kim, Y. (2014, October). Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1746-1751).

- [34] Gu, X., Gu, Y., & Wu, H. (2017). Cascaded convolutional neural networks for aspect-based opinion summary. *Neural Processing Letters*, 46(2), 581-594.
- [35] Du, H., Xu, X., Cheng, X., Wu, D., Liu, Y., & Yu, Z. (2016, April). Aspect-specific sentimental word embedding for sentiment analysis of online reviews. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 29-30).
- [36] Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016, August). Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 207-212).
- [37] Yang, M., Tu, W., Wang, J., Xu, F., & Chen, X. (2017, February). Attention based LSTM for target dependent sentiment classification. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [38] Ma, Y., Peng, H., & Cambria, E. (2018, April). Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In *Thirty-second AAAI conference on artificial intelligence*.
- [39] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016, June). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1480-1489).
- [40] Li, L., Liu, Y., & Zhou, A. (2018, October). Hierarchical attention based position-aware network for aspect-level sentiment analysis. In *Proceedings of the 22nd Conference on Computational Natural Language Learning* (pp. 181-189).

- [41] Zhu, X. (2007, June). Semi-supervised learning tutorial. In International Conference on Machine Learning (ICML) (pp. 1-135).
- [42] Srivastava Aman, Reddy SL Ruthvik (2020). Spello <https://github.com/hellohaptik/spello>
- [43] Popescu, A. M., & Etzioni, O. (2005). Extracting Product Features and Opinions from Reviews. In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pages 339-346. Vancouver, British Columbia, Canada.
- [44] Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In Advances in neural information processing systems (pp. 649-657).
- [45] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).
- [46] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.
- [47] Han, H., Wang, W. Y., & Mao, B. H. (2005, August). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In International conference on intelligent computing (pp. 878-887). Springer, Berlin, Heidelberg.

Abstract

Aspect-Level Analysis and Predictive Modeling for Electric Vehicle Based on Aspect-Based Sentiment Analysis Using Machine Learning

Seijun Chung

Department of Industrial Engineering
The Graduate School of Engineering
Seoul National University

In this study, we extract main components and attributes, which are the main aspects of Electric Vehicle by analyzing User Experience based on Aspect-Based Sentiment Analysis (ABSA) using machine learning, overcoming the problems accompanying in this process such as Data Imbalance and insufficient user reviews by making use of non-label data. In addition, we find the contributing factors affecting user's sentiments by figuring out the relationship between user's sentiment to each aspect extracted and detailed specifications of Electric Vehicle with regression.

Based on the ABSA method, and we perform data collection, data preprocessing, feature engineering, Aspect Extraction, modeling for sentiment analysis, and evaluating user sentiment to each aspect in sequence. For data collection, a total of 5,065 label data, which is evaluated with a 5-point scale by users, was collected from representative car forums. At the same time, in order to overcome the shortage of data and data imbalance, approximately 210,000 items of non-label data are collected from Youtube.com, of which 6,488 items were selected by filtering with limited to the user experience related only. And then, feature engineering is performed with effective embedding methods of distributed representation after data pre-processing. The analysis phase is mainly divided into two processes: Aspect

Extraction and Sentiment Analysis. First of all, TextRank and Naïve methods were used as an unsupervised method and an extractive approach for Aspect Extraction. Then, in order to implement a sentiment classification model based on supervised learning with high performance, we built a machine learning model that trains the truncated text composed of one or two sentences at the beginning of a review text with a label and make it improved by means of semi-supervised learning. With the model trained, we are able to perform aspect-wise sentiment analysis by conducting sentiment analysis on the sentence that including the selected aspect term. Further, we find detailed specifications of vehicle that have an influence on user sentiment as contributing factors that affects user's sentiment.

As a result, 16 categories of main aspects were extracted, eight key EV Components & eight key Human Factor Attributes, of which the users are likely to be positive to "Acceleration, Room, Interior, Power, Safety, Ergonomics, Price, Power" and negative to "Seat, Battery, Charge, Noise, Winter, Ice". In sentiment analysis, the CNN model showed the highest performance in sentiment classification. Therefore, through semi-supervised learning using CNN, label propagation was performed among non-label data, giving the pseudo label to only the data with a high classification probability more than 80%, resulting in improvement in performance of the CNN model. Lastly, we confirmed the high classification accuracy of the deep learning model for predicting the user's sentiment of the sentences. In addition, with regard to aspect-wise sentiment analysis, there was a tendency to predict the user's sentiment similarly between machine learning based and lexicon-based, which showed machine learning based model is robust as much as lexicon-based.

In conclusion, it was shown that more diverse topics and unbiased opinions could be extracted through aspect-wise analysis than review-wise. In addition, we verified that the imbalance problem could be overcome by over-sampling. Finally, a more effective UX analysis framework for the products that have not sufficient user reviews was proposed by taking advantage of non-label data with semi-supervised learning.

Keywords: Aspect Based Sentiment Analysis, User Experience, Machine Learning

Student Number: 2018-23341

감사의 글

저에게 새로운 학문의 길을 열어주신 윤명환 교수님, 그리고 늦깎이 대학원생인 저를 빠르게 연구에 적응할 수 있도록 많은 도움을 주신 휴먼인터페이스시스템 연구실의 모든 소속원들께 감사의 인사 올립니다.

그리고 제가 대학원에 진학하여 학문에 전념할 수 있도록 응원해주신 사랑하는 우리 가족, 아버지, 어머니, 장인어른, 장모님 그리고 저를 끝까지 믿고 지지해준 영원한 동반자 헤림이에게 감사하다는 말씀 전하고 싶습니다. 앞으로도 기대에 부응할 수 있도록 더욱 열심히 정진하는 삶을 살도록 하겠습니다.

2020년 8월

정 세 준